Econometric Analysis of the Real Exchange Rate, Energy Consumption and Instrumental Variables.

Anna Brocklebank

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Social Sciences
Accounting, Economics and Finance

April 2018

Abstract

The present thesis consists of four chapters: the introduction (Chapter 1), and three econometrics-based research papers (Chapters 2 through 4). The introduction sets out the research questions explored in the subsequent chapters and previews the results.

Chapter 2 uses Monte Carlo studies to investigate a pre-test bias problem potentially associated with using Stock and Yogo's (2005) critical values to determine whether the instrumental variables in the model are strong or weak. A solution is proposed based on Angrist and Krueger's (1995) Split-Sample IV estimator.

Chapter 3 tests the Balassa-Samuelson hypothesis using a new dataset, and with a focus on oil-producing countries. A relationship between oil rents, which represent productivity measure in the tradable sector of oil exporters, and the real exchange rates are examined using unique data constructed from Wood Mackenzie's Global Economic Model. The results provide evidence in favour of the Balassa-Samuelson effect in most countries in the sample, apart from OPEC countries.

Chapter 4 investigates the long-run relationship between GDP per capita, energy consumption and energy prices in a set of 28 OECD countries using newer econometric techniques than have been prevalent in the prior literature. The results of the analysis suggest that the long-run bi-directional relationship exists, but is likely to be heterogeneous across countries. Also, the bi-directionality is not symmetric: energy consumption was found to be more strongly affected by economic growth than vice versa.

In a narrative sense, the three research papers reflect the progression of my interests over the past several years: from a work focused on econometric theory ('how should we do IV pre-testing?'), to a work of applied econometrics testing a piece of economic theory ('does the Balassa-Samuelson hypothesis account for real exchange rate movements in oil exporters?'), and finally to a work of applied econometrics which is oriented more towards answering policy-relevant questions ('if we cut our energy usage to meet climate targets, will we choke off economic growth?').

Acknowledgements

First of all I would like to thank my supervisor, Professor Mark Schaffer, for the tremendous help provided throughout these many years of PhD study interlaced with a couple of maternity breaks. His comments have been invaluable, and I have learned so much from him in a whole variety of settings (research, teaching, BP’s Annual Review, et cetera). I can hardly imagine what my PhD would have been like without Mark. I’m also very grateful to my husband, Sean Brocklebank, for his willingness to have long conversations about my research and all the suggestions and recommendations he provided (he also helped me convince Wood Mackenzie to hand over the data for Chapter 2). I also want to thank my second supervisor, Dr. Atanas Christev, for all the support I got from him. And Erkal Ersoy for the work we have done together. Finally, my kids and my parents: the kids for sleeping through (most) nights, and my parents for being always there whenever I needed them.

Declaration.

Chapter 3 is joint work with Erkal Ersoy (a fellow Economics PhD student at Heriot-Watt), with my contribution being 55-60% of the work. A big part of the work was data clean-up and aggregation (Wood Mackenzie collected the data by oil rig and we needed to transform it to the country-level). Erkal and I split all oil rigs in equal proportion. The main body of the analysis we did together. After the joint draft was formed, I have made miscellaneous changes to all the sections, apart from the data description section. I also added section 3.6.4 «Further investigation of OPEC countries» to the analysis, and wrote the conclusion.

Research thesis submission form should be placed here

Contents

Chapter 1. Introduction

“A good sketch is better than a long speech.”
– *Napoleon Bonaparte*

Taking a cue from the French Emperor, this introduction takes a graphical view of the whole thesis and shows that, chapter by chapter, the results have an important thing in common – they can all be illustrated with a few relatively simple charts. So in the course of the next few pages, I attempt to convince the reader that: (i) pre-testing for weak instruments will often introduce bias (but using split-sample techniques can solve this problem), (ii) the relationship between the real exchange rate and the productivity of the oil sector in the world’s biggest oil exporters – which you would expect to be strongly positive on the basis of the Balassa-Samuelson effect – is actually pretty weak, and (iii) within the G7 countries, there is a relatively consistent causal relationship from GDP to energy usage, but not from energy usage to GDP (the picture for non-G7 OECD countries is more mixed).

Second Chapter: ***Should we pre-test instrumental variables? A Monte Carlo study***

Weak instruments are a pervasive issue in the microeconometric literature, and pre-tests such as those based on Stock and Yogo’s (2005) first stage F-statistic critical values have been widely used to assess whether an instrument is too weak or if it is strong enough. But there is a problem – if the instruments are weak (i.e. weakly correlated with the endogenous regressors in the model), then the sampling distribution of the Two-Stage Least Squares (2SLS) estimator doesn’t follow the normal distribution, and any subsequent inference is unreliable. It is this problem of unreliable inference which is the subject of this chapter.

Stock and Yogo (2005) suggested that strength of the instrument should be viewed in terms of the maximal relative bias of the IV estimator and the maximal size of the Wald test when performing hypothesis testing on the parameter of interest. I set up an experiment based on Hall et al.’s (1996) simulation and investigate how Stock and Yogo’s critical values perform in practice. I look at the simulated samples when the first stage F-statistic exceeds the critical value and compare the observed Wald test size distortions to those predicted by the critical values. You can get an idea of these results in Figure 1.1, which shows the size distortions to the Wald test as a function of the strength of the instrument (the correlation between an endogenous regressor, x, and an instrument, z, – increasing from 0 to 0.3 across the clusters of bars), and as a function of the degree of endogeneity (the correlation between x and u – increasing from 0.1 to 0.9 within each cluster of bars). The size of the distortions (the actual size of the test with the nominal size of 5%) for all samples are shown in red, while the size distortions for those values which exceed Stock and Yogo's critical values are shown in blue. Three things stand out in the picture: (i) the distortions are larger when the instrument is weaker (as in the leftmost clusters), (ii) the distortions are larger when the degree of endogeneity is greater (as is the case in the rightmost values within each cluster), and (iii) the distortions are much larger for those samples where the F-statistic exceeded Stock and Yogo's critical values than for the whole sample.

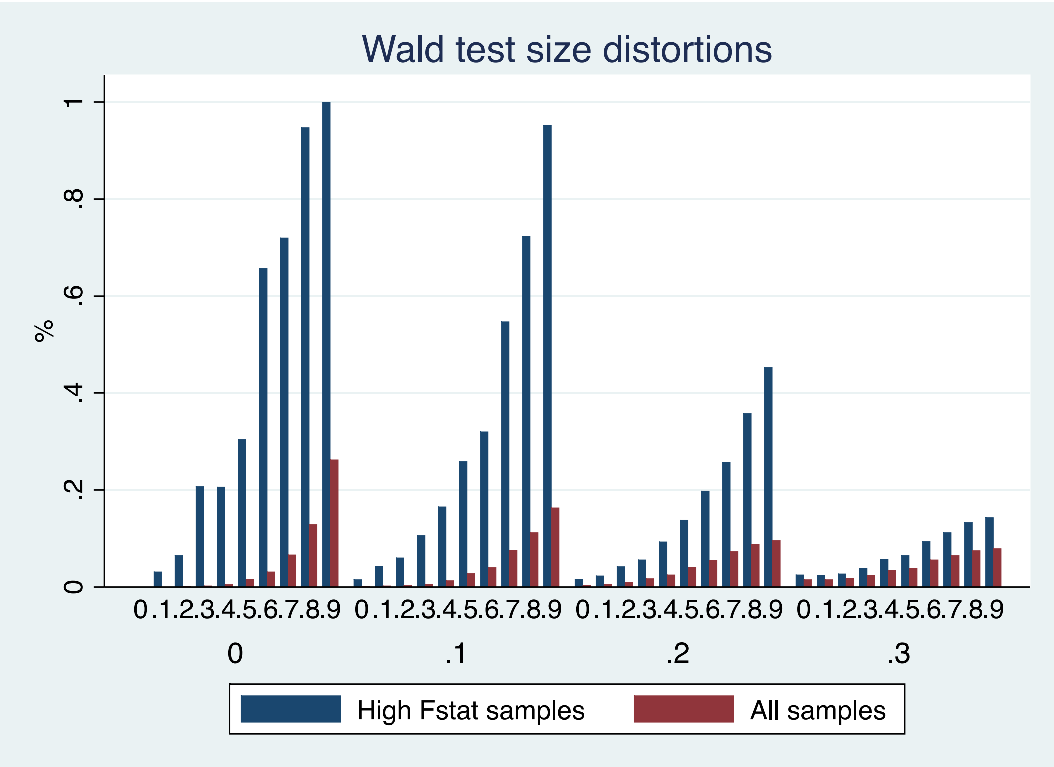

Figure 1.1 Wald Test Size Distortions. The larger numbers at the base ranging from 0 to 0.3 represent the strength of the instrument (the correlation between x and z), while the small repeated numbers ranging from 0.1 to 0.9 under each set of bars represent the degree of endogeneity (the correlation between x and the error term u). The nominal size of the test is 5%.

Another contribution of this work is to use a median-squared error as a loss function to evaluate the effectiveness of the pre-test. The result of taking this view of the data is shown in Figure 1.2. Two things stand out in this second picture: (i) the distortions are larger when the instrument is weaker (as in the leftmost clusters), and (ii) the distortions are larger when the degree of endogeneity is greater (as is the case in the rightmost values within each cluster), but only for the samples with the largest F-statistics.

Taking both views from 1.1 and 1.2 together, using Stock and Yogo's critical values as a pre-test is most problematic when the the degree of endogeneity is high (i.e. when instruments are most necessary) and when the instruments are weak (i.e. much of the time).

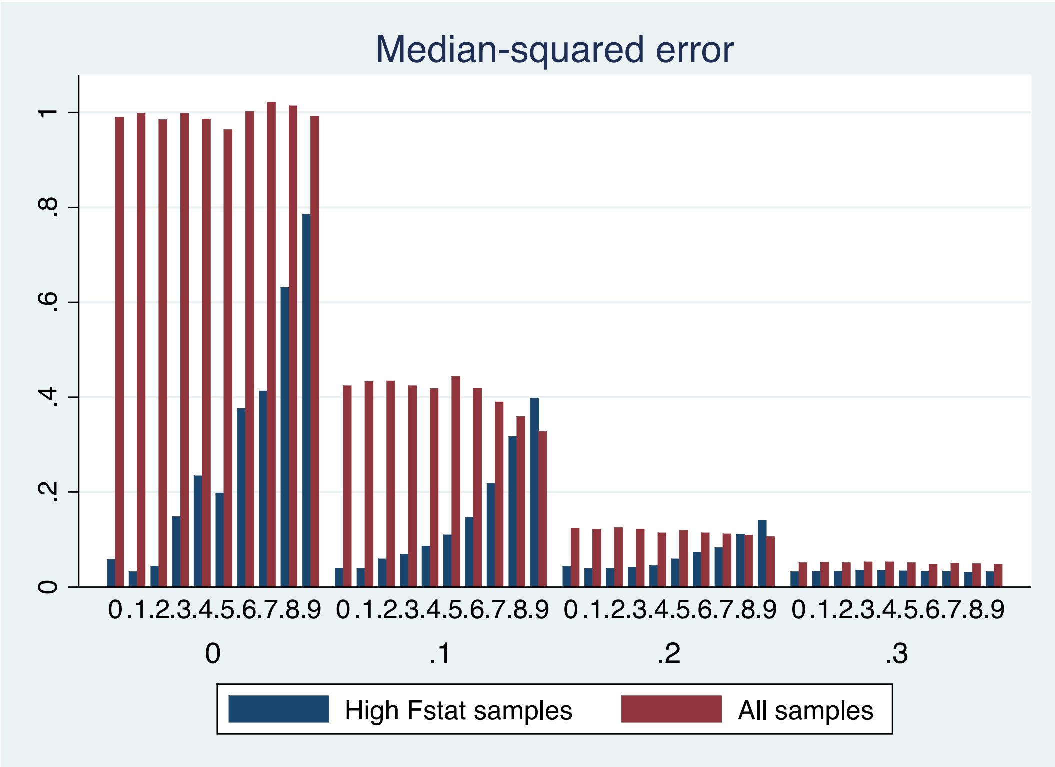

Figure 1.2. Median-squared error. The larger numbers at the base ranging from 0 to 0.3 represent the strength of the instrument (the correlation between x and z), while the small repeated numbers ranging from 0.1 to 0.9 under each set of bars represent the degree of endogeneity (the correlation between x and u).

So instrument selection based on the instrument "passing" the pre-test is associated with a potential pre-test bias. At the same time, I find that when the degree of endogeneity of the regressor is low, choosing an instrument which is associated with a higher first stage F-statistic could improve the precision of estimation of the parameter of interest. You can see this in Figure 1.2 by noting that the "high F-statistic" samples on the left within each cluster have lower median-squared error than the sample as a whole.

So what practical steps can a researcher take when they are dealing with weak instruments? I suggest a solution to the pre-test bias problem based on Angrist and Krueger's (1995) Split-Sample IV estimator (SSIV). As the name suggests, the basic idea is to split the sample into two parts: one part to test the strength of the instruments, and the other part to estimate the coefficients of interest. As you can see by contrasting Figures 1.3 and 1.4 (both graphs focus on "high F-statistic" samples), the use of this procedure largely eliminates the pre-test bias problem. The size distortions are still increasing in the degree of endogeneity in Figure 1.4, but much less so than in Figure 1.3, which uses the standard 2SLS procedure. The use of SSIV also means that median-squared error is not increasing in the degree of endogeneity (in contrast to the results under 2SLS). However, the accuracy of estimation of the parameter of interest suffers notably, which suggests that the approach can only be justified when the sample size is large.

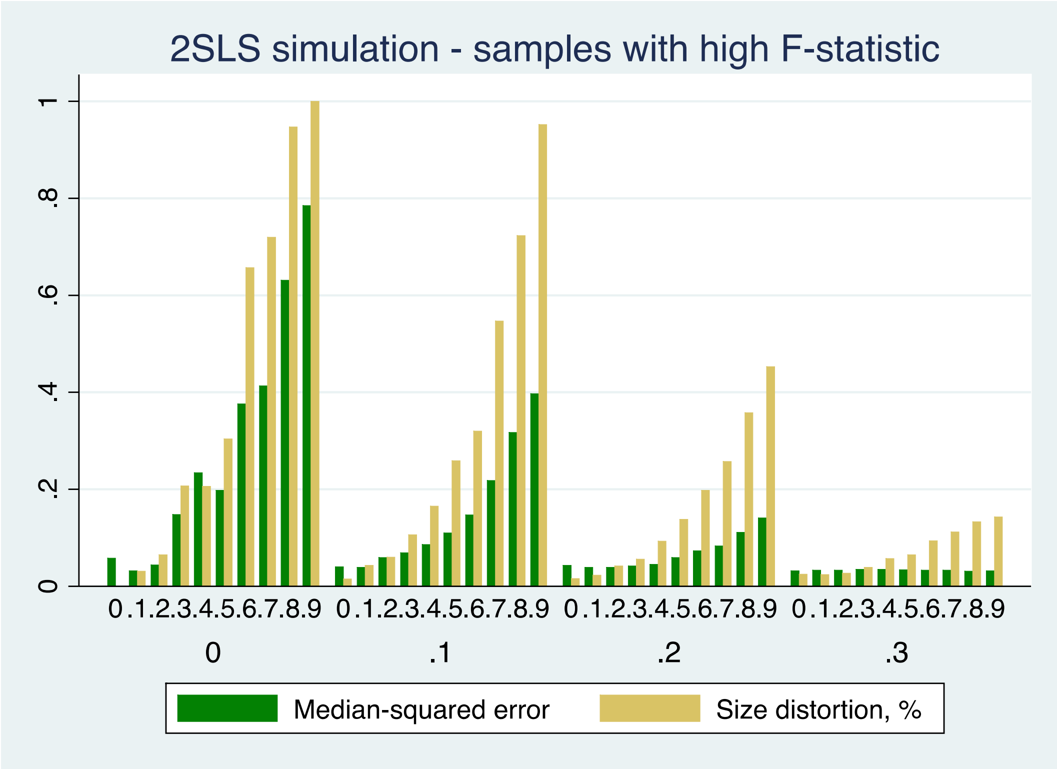

Figure 1.3 A Simulation of the Standard Two-Stage Least Squares Approach with Stock and Yogo's critical values. The larger numbers at the base ranging from 0 to 0.3 represent the strength of the instrument (the correlation between x and z), while the small repeated numbers ranging from 0.1 to 0.9 under each set of bars represent the degree of endogeneity (the correlation between x and u).

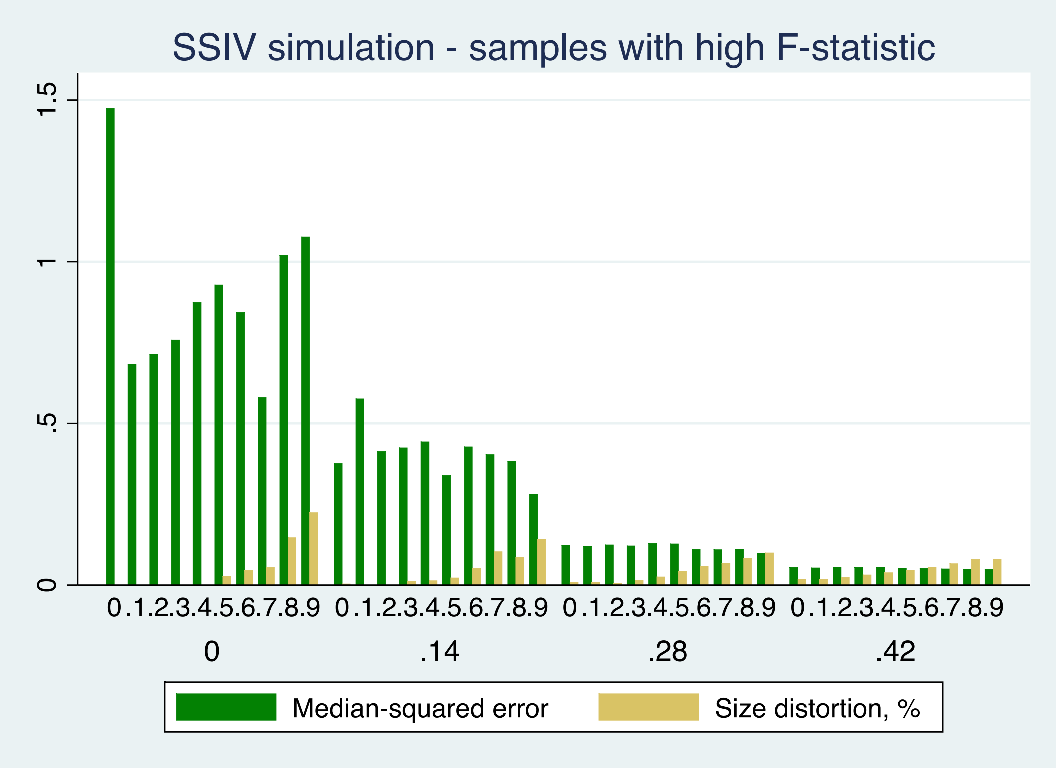

Figure 1.4 A Simulation of the Split-Sample Instrumental Variables Approach. The larger numbers at the base ranging from 0 to 0.42 represent the strength of the instrument (the correlation between x and z), while the small repeated numbers ranging from 0.1 to 0.9 under each set of bars represent the degree of endogeneity (the correlation between x and u). The sample is split evenly into two halves.

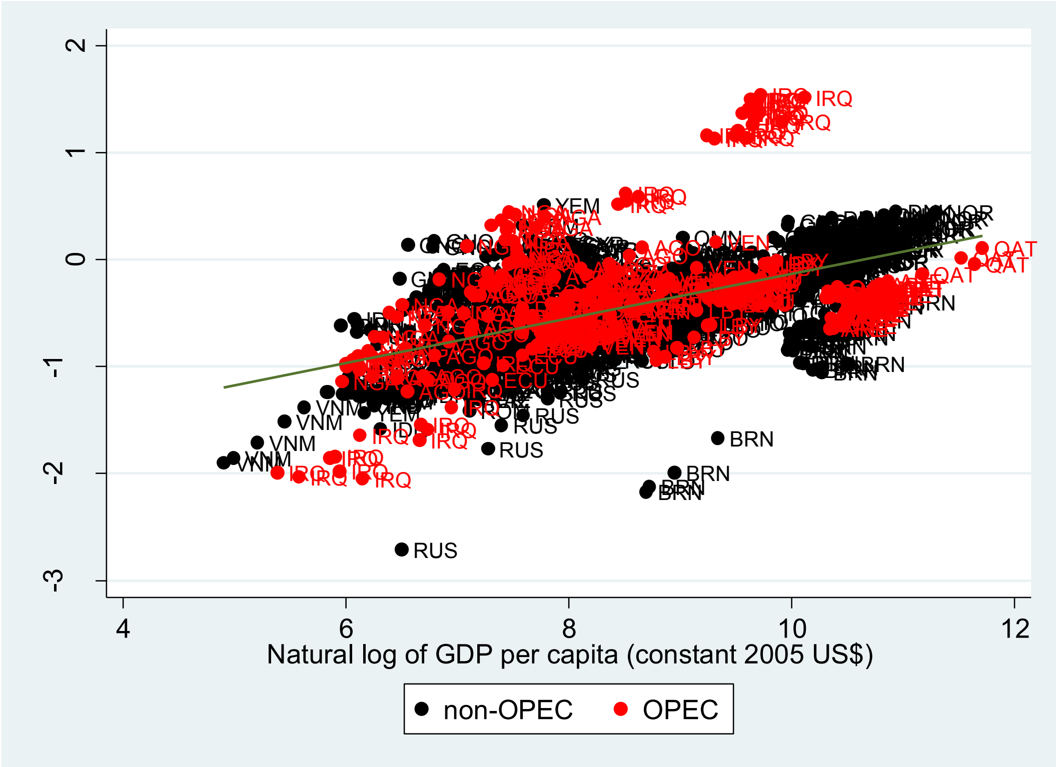

Figure 1.5 The Penn Effect. The price level (measured relative to the United States), shown on the vertical axis, tends to be higher the higher is GDP per capita. The relationship holds both in OPEC and non-OPEC countries. The data cover 42 countries over the period 1965-2009. Source: Penn World Table (PWT 7.1)

The Penn Effect is an empirical regularity – price levels tend to be higher in wealthier countries. You can see this clearly in figure 1.5. The Balassa-Samuelson hypothesis is an explanation for this regularity: when country's tradable sector becomes more productive and experiences wage growth, country's non-tradable sector will experience wage growth as well, as long as workers can freely move between sectors. The increase in wages in both non-tradable and tradable sectors will drive the price level up, which in turn will lead to the real exchange rate appreciation and higher real price levels.

In this chapter, we exploit the proprietary data, obtained from Wood Mackenzie's (WM) Global Economic Model (GEM), which allowed us to construct oil production costs and

[1] Joint work with Erkal Ersoy

revenues for a number of oil-exporting countries. Based on these costs and revenues, oil rents (i.e. profits) are calculated. As oil rents represent productivity in the oil sector, we use them to test for evidence of the Balassa-Samuelson hypothesis in oil-exporting countries. We specifically focus on testing whether the increase in the productivity of the oil sector leads to real exchange rate appreciation of the oil exporters.

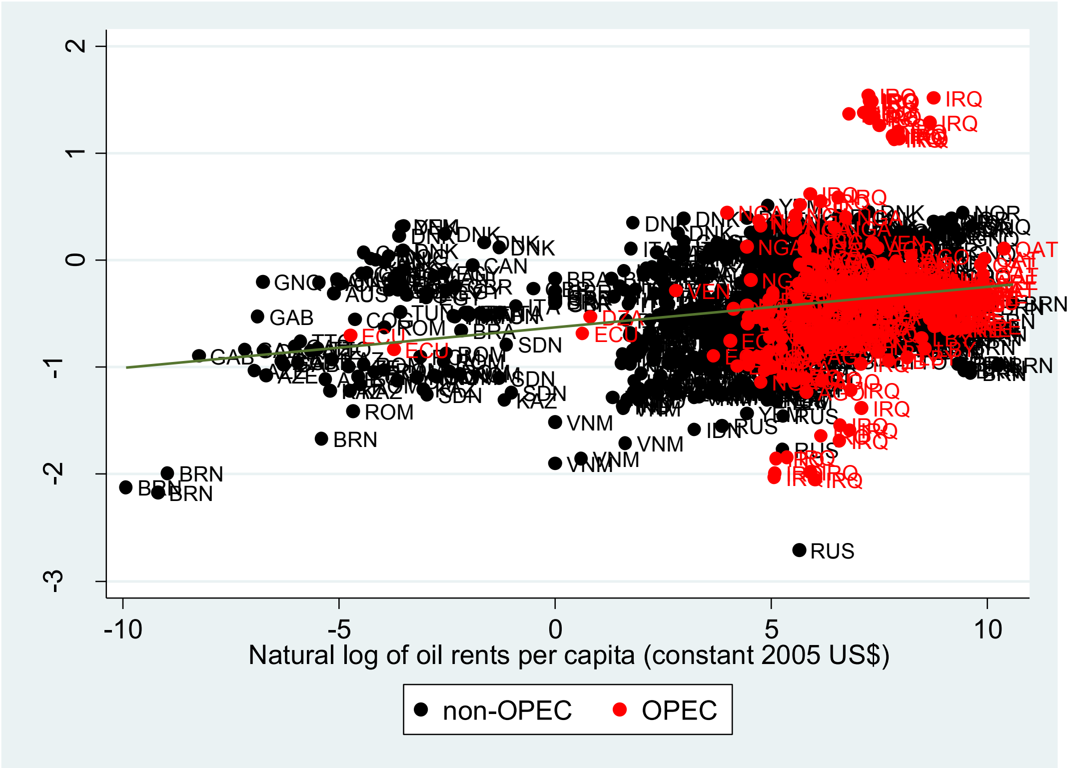

Figure 1.6 Testing Balassa-Samuelson. The price level (measured relative to the United States), shown on the vertical axis, tends to be higher the higher is productivity in the oil sector (measured as rents per capita). Surprisingly, however, this relationship does not seem to hold in OPEC countries (shown in red). . The data cover 42 countries over the period 1965-2009. Source: Penn World Table (PWT 7.1), Wood Mackenzie's (WM) Global Economic Model (GEM)

The results we obtained were unexpected – they suggested that while for most countries in our sample of 42, the Balassa-Samuelson mechanism does hold, there was no evidence for the hypothesis in OPEC (Organization of the Petroleum Exporting Countries). In OPEC, neither the nominal exchange rate nor the price level were affected by an increase in oil sector productivity. We attribute the failure of the mechanism to the fact that at least one of the assumptions of the Balassa-Samuelson hypothesis – free movement of labour between tradable and non-tradable sectors – is likely to be violated. For the rest of the sample, we find that the effect of oil productivity, while significant, is very small - about 8 times smaller than the effect of the oil price movements.

Fourth Chapter: ***How Costly is Conservation? The International Energy-GDP Relationship Re-examined***

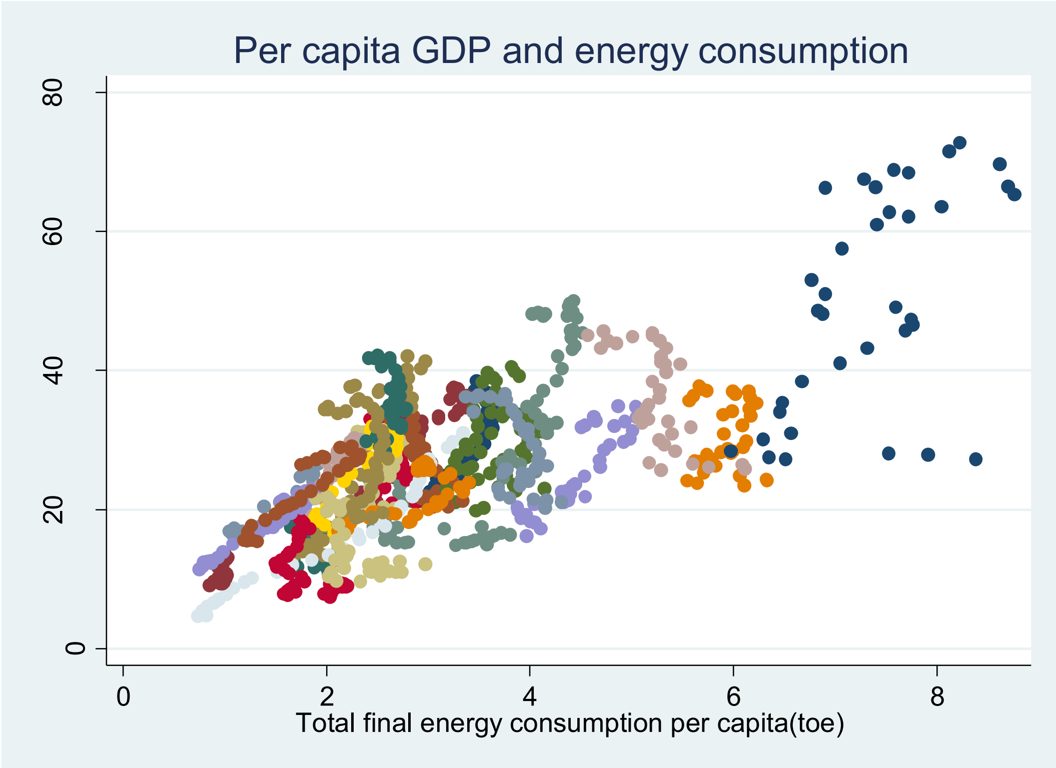

Figure 1.7 GDP per capita and energy consumption. Comparing total energy consumption (measured in tons-of-oil-equivalent) to PPP GDP per capita, we can see that there is a positive relationship. In this graph, each point represents one year of data for one country, and country is its own colour (the full set of countries is listed in the chapter).

There is a positive correlation between economic growth and energy consumption, and we can see that clearly in Figure 1.7 (note: in the picture each country and year pair is a single dot, and each country has its own colour). But what is the causal relationship? Does economic growth cause increased energy consumption? Does increased energy usage lead to economic expansion? Both? Neither? This question matters as we look forward to a future where energy use may end up being distinctly lower in order for us to meet climate change goals. So this chapter reviews the long-run relationship between economic growth, energy consumption and energy prices for 28 OECD (Organisation of Economic Cooperation and Development) nations and for the subset of G7 countries. Despite decades of research in this area, there has been no agreement on whether economic growth drives energy consumption or the other way around; the other two possibilities are that causality is bidirectional or the two are independent. The earlier studies used country-level data and time series estimators to analyse the relationship (Masih and Masih, 1997; Stern, 2000) . Newer studies (for example Belke et al., 2011; Damette and Seghir, 2013). take advantage of available panel datasets and panel estimators, which offer efficiency gains compared to individual country-level analysis.

Throughout the chapter I employ most up-to-date econometrics techniques to first test the order of integration of the variables (Bai and Ng, 2004, 2010), then to test if the long-run relationship exists (Westerlund's 2007 cointegration test) and lastly to estimate parameters of the long-run relationship using panel Dynamic Ordinary Least Squares estimator by Mark and Sul (2003), Mean Group (Pesaran and Smith, 1995), Pooled Mean Group (Pesaran et al., 1999) and Common Correlated Effects Mean Group (Pesaran, 2006) estimators. I compare the results of the Mean Group estimations with the results produced by other panel estimators, and I show that the relationship between energy consumption and GDP per capita is most likely heterogeneous across countries. I argue that panel estimators have to be chosen carefully to allow for this heterogeneity. I use Pesaran's (2015) test to show that cross-sectional dependence is likely to be an issue and I re-estimate the relationship using CCEMG which is consistent in its presence.

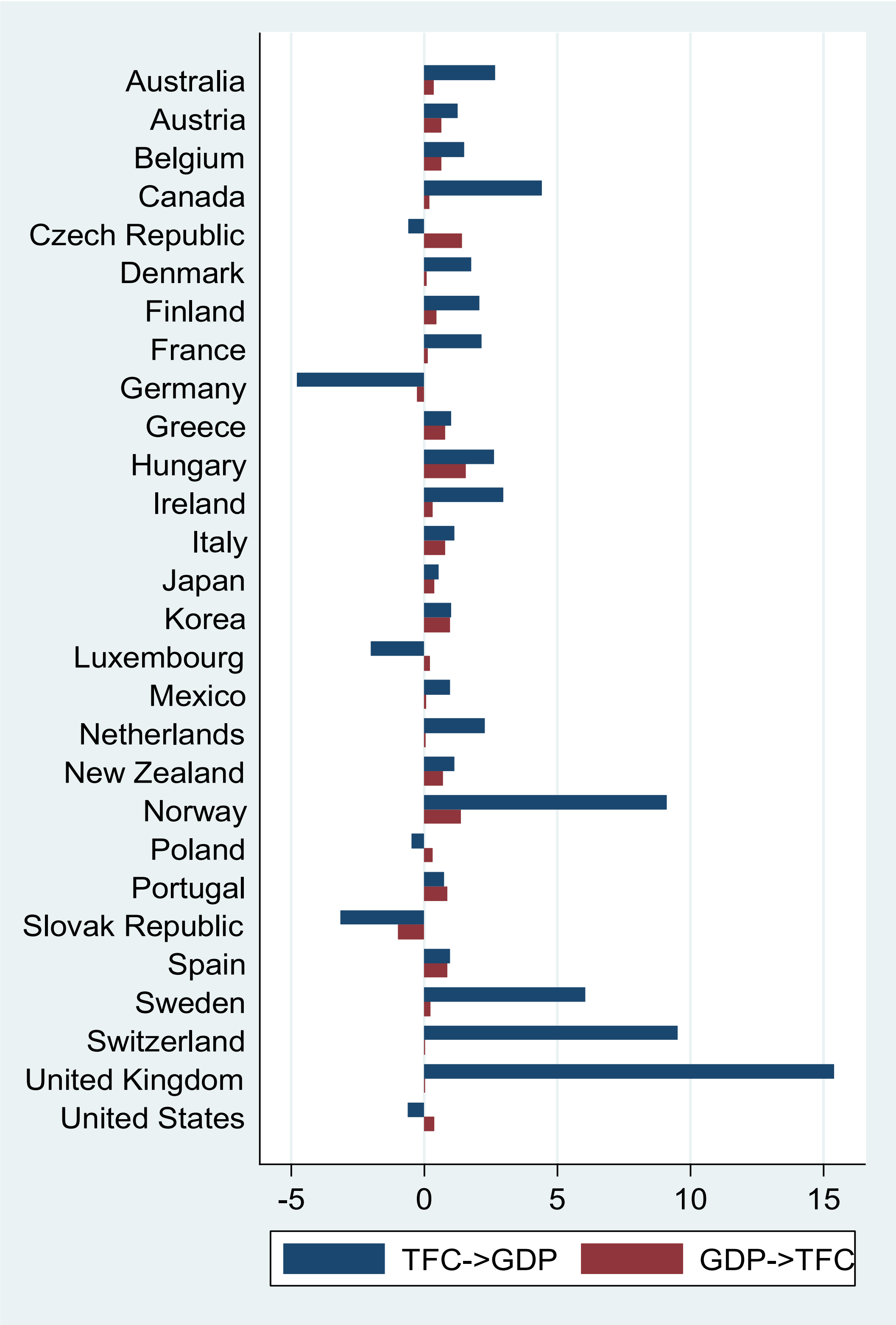

Figure 1.8 Estimated effects of TFC (Total Final energy Consumption, toe) per capita on GDP per capita (blue) and of GDP per capita on TFC (red). In each case, the effect is measured as an elasticity. Notice that while the GDP -> TFC elasticities are relatively stable, plausibly sized and almost all have the expected sign, the TFC -> GDP elasticities are unstable, often implausibly large, and frequently negative.

The details of the estimation are explained fully in the chapter, but again the basic results can be summarised in a graph, in this case Figure 1.8. Although the main results were estimated as a panel (which increases statistical power), this figure depicts the country-by-country estimates of elasticities for the two causal channels – from energy consumption to GDP (in blue) and from GDP to energy consumption (in red). We can see right away that the blue, energy-to-GDP estimates are unstable and frequently implausibly large or implausibly negative. The red, GDP-to-energy elasticity estimates, in contrast, are both more stable across countries and more realistic in magnitude. So while these country-by-country estimates were not the basis for inference in the main chapter, they do reflect the main finding, which is that the causality is essentially one-directional: economic growth leads to increased energy consumption. This one-directional effect is significant and relatively similar across countries. There is some weak evidence for causality running from energy consumption to GDP and the relationship appears to be heterogeneous across countries. Moreover, there was no causal effect found in the G7 countries. These results suggest that modest energy conservation policies are not likely to have a serious negative effect on economic growth in the long run.

To summarise, the three core chapters reflect the progression of my interests over the past several years: from a work focused on econometric theory ('how should we do IV pre-testing?'), to a work of applied econometrics testing a piece of economic theory ('does the Balassa-Samuelson hypothesis account for real exchange rate movements in oil exporters?'), and finally to a work of applied econometrics which is oriented more towards answering policy-relevant questions ('if we cut our energy usage to meet climate targets, will we choke off economic growth?').

Chapter 2. Should we pre-test instrumental variables? A Monte Carlo study.

2.1 Introduction.

Instrumental variables techniques are widely used in applied research, particularly because they allow economists to distinguish and evaluate causal relationships in the absence of controlled experiments or to estimate equations where the regressor is unobservable or measured with error. In order to proceed with an instrumental variables (IV) estimation, the researcher has to find appropriate instruments, which is often quite difficult. Once some instruments are found which satisfy the necessary conditions, i.e. instrument exogeneity and instrument relevance, they may still compromise the reliability of inferences on the parameters of interest if the instruments are not sufficiently correlated with the endogenous regressors. The question of whether the instruments at hand are strong enough to provide reliable test statistics is not easy to answer. Stock and Yogo (2005) derived critical values for the first stage F-statistic in the IV regression that enable researchers to determine whether the instruments used are strong or weak; the strength of the instruments is defined in terms of the maximal bias of the IV estimator relative to the OLS estimator and the maximal size distortions of the Wald test while testing hypotheses on the coefficient of the endogenous regressors.

Stock and Yogo's critical values are routinely reported by Stata's ivreg2 command (Baum et al., 2007) and widely used in applied work (see for example Amiti and Konings (2007), or Kilian (2008)), because they seem to have provided a solution to a serious problem that researchers face - how to determine whether the estimation results based on IV methods are trustworthy or not. In this paper, I draw attention to a potential problem associated with the use of Stock and Yogo's critical values as a pre-test to determine whether the instruments are strong or weak. The problem is that choosing the instruments based on a pre-test and then estimating the main equation using preselected instruments introduces problems associated with sequential testing. Depending on how the pre-test is carried out a bias might be introduced, something that researchers should be cautious about.

Another innovative contribution of my paper is in using median-squared error as a loss function in place of more traditional mean-squared error, which can not be applied if the model is exactly identified. On the basis of the the loss function, the effectiveness of the pre-test is evaluated.

I also suggest a solution to the problem associated with the pre-test bias, which is based on Angrist and Krueger's (1995) Split-Sample IV estimator (SSIV). I provide the results of Monte Carlo simulations that support both claims I will make in this paper – (i) that we should be cautious when using pre-tests and (ii) that SSIV might be a useful alternative to Two-Stage Least Squares estimator (2SLS) in the context of weak IV detection, instrument selection, and estimation of the main equation of interest.

The chapter is structured as follows: section 2.2 outlines the model and its assumptions; section 2.3 considers the consequences of having weak instruments in the IV regression; section 2.4 discusses detection of the weak IV problem, in particular the test suggested by Stock and Yogo (2005) which is based on the critical values for the first stage F-statistic; section 2.5 considers the potential problem that comes from pre-testing is reviewed; section 2.6 describes the design of the Monte Carlo simulations; section 2.7 investigates the performance of the critical values based on the results of the simulations; sections 2.8 and 2.9 point out caveats with which the critical values should be used in applied work and explore an alternative way of using the critical values in order to overcome the problem of sequential testing; section 2.10 concludes the analysis.

2.2 Model and assumptions.

Consider the linear IV model with one endogenous regressor and no exogenous regressors:

$$Y = X\beta + \varepsilon \tag{2.1}$$

where Y is an $N \times 1$ vector of observations on the dependent variable, X is an $N \times 1$ vector of observations on the endogenous regressor, ε is an $N \times 1$ vector of the iid error term distributed normally N(0, σ_u^2) and β is a scalar parameter of interest. This equation is referred to as the structural equation.

In row notation:

$$y_i = x_i\beta + \varepsilon_i \tag{2.2}$$

The reduced form equation for X:

$$X = Z\Pi + v \tag{2.3}$$

where Z is an $N \times L$ matrix of instruments, v is an $N \times 1$ vector of the iid error term distributed N(0, σ_v^2) and Π is a $L \times 1$ vector of unknown coefficients; corr$(\varepsilon_i, v_i) = \rho$.

In row notation

$$x_i = \boldsymbol{z}_i\pi + v_i \tag{2.4}$$

The instruments are assumed to be correlated with the endogenous regressor X, $rank(\boldsymbol{z_i}'x_i) \neq 0$ and uncorrelated with the error term ε, $E(z_i\varepsilon_i) \neq 0$. The former is referred to as the rank conditions and says that instruments should be relevant, i.e., correlated with the endogenous regressors; the latter says that instruments should be weakly exogenous.

The most commonly used IV method in applied work is the Two-Stage Least Square estimator (2SLS):

$$\beta_{2SLS} = (X'P_zX)^{-1}(X'P_zY) \tag{2.5}$$

where $P_z = Z(Z'Z)^{-1}Z'$ is a projection matrix.

The 2SLS estimator is usually thought of as a two-step estimator, where the first step is based on the reduced form equation, and the second step is based on the structural equation, but instead of regressing y on x, y is regressed on $\hat{x}$, where $\hat{x}$ is a predicted value from the first step regression.

In order to determine whether the instruments are satisfying the non-zero rank condition, the researcher needs to perform an underidentification test – a test of the rank of the Π matrix. The model considered in this paper only has one endogenous regressor, so Π is a vector, and, as noted in Baum et al. (2003), the underidentification test is equivalent to testing whether the instruments are jointly significant at explaining x. If the F-statistics from the first stage regression exceeds the critical value, then the Ho of Π=0 is rejected at the corresponding significance level.

Testing weak exogeneity can prove to be a harder task – in order to perform it the researcher would need to have several instruments at hand and would have to have prior beliefs or theories on which instruments are more reliable and informative. This paper will not focus directly on the issues related to the failure of the weak exogeneity condition for the instruments.

Weak exogeneity and the non-zero rank condition are crucial for the validity of the IV regression: they ensure that the IV estimator is identified and consistent. However, even if these two conditions are satisfied, this doesn't necessarily imply reliable inference – when the instruments are correlated only weakly with the endogenous regressors, there can be a large bias in the IV estimator and size distortions of the test of the parameter of interest in the structural equation.

2.3 Sample size vs. weak IV problem.

When the rank condition and weak exogeneity are satisfied, the IV estimator is consistent, but still biased on the way to infinity. When N is not sufficiently large, sampling distribution of the 2SLS is not well approximated by the normal distribution, the confidence intervals produced are incorrect and the inference is unreliable. However, Bound, Jaeger and Baker (1995) showed that in the presence of weak instruments even when the sample size is large, IV estimator still suffers from all the aforementioned problems and the weaker the correlation between Z and X is the more severe are the distortions of the sampling distributions of the statistics.

Nelson and Startz (1990b) analyzed the small sample distribution of the IV estimator in the case with one endogenous regressor and one instrument by deriving its pdf and cdf and investigating their behavior for different values of the parameters. They concluded that the distribution is very different from its asymptotic approximation. In fact, they found that the distribution is bimodal and the center of it is closer to the probability limit of the OLS estimator and not the true value of zero. They noted that the approximation was particularly poor when the sample size was small or the instrument was weak.

Why do weak instruments and small samples seem to have similar effect on the distribution of the IV estimator? The answer to this question can be found in Rothenberg (1984), where he discusses the so-called concentration parameter μ^2:

$$\mu^2 = \frac{\Pi' Z' Z \Pi}{\sigma_v^2} \tag{2.6}$$

He expressed the 2SLS estimator in terms of the standardized normal variables and μ, which allowed him to show that the sample size affects the distribution of 2SLS estimator only through the concentration parameter and when μ^2 is large, the standardized 2SLS estimator is distributed as $N(0,1)$. The concentration parameter will usually be large when the sample size is large, but it can also be large if σ_v^2 is small. Conversely, even if the sample size is large, the concentration parameter can be small if the σ_v^2 is also large, which would be the case if significant proportion of the variation in X stays unexplained by the instruments, or in other words instruments are weakly correlated with the endogenous regressor. Rothenberg (1984) also suggests that it would be more useful to scale the distribution of the estimator in terms of $\frac{1}{\mu}$ rather than $\frac{1}{\sqrt{n}}$ in the context of asymptotic analysis.

Stock et al. (2002) provide graphical illustration of the relationship between μ^2 and the probability density function (pdf) of the 2SLS estimator and its t-statistic, which is based on the model used by Nelson and Startz (1990b). According to Stock's et al. (2002) Monte Carlo simulations for the one instrument case with the correlation between the error terms in the first stage regression and the structural equation at $\rho = 0.99$, the distribution of the 2SLS estimator is centered around the same wrong value as the OLS estimator when $\frac{\mu^2}{L} = 0$, when $\frac{\mu^2}{L}$ is larger than zero but small, the distribution is bimodal, and finally (only) when $\frac{\mu^2}{L} > 10$, will the distribution approach its asymptotic approximation. The graphs also show the pdfs of the corresponding t-statistics; when μ^2 is small, the distributions are heavily skewed to the right, which would cause large distortions to the size of the t-test.

In their earlier research, Staiger and Stock (1997) introduced an alternative approach to approximating distributions in the presence of weak instruments. Unlike the traditional approach that assumed fixed, nonzero coefficients in the first stage of the IV regression, Staiger and Stock (1997) suggested fixing the first stage F-statistic by lowering the value of the first stage coefficients as the sample size increases. Contrary to conventional asymptotics, the F-statistic under the the alternative approach does not tend to infinity as the number of observations goes up. The new approach enabled the authors to derive asymptotic representations of multiple test statistics, which was not possible using the traditional approach.

One example of the sort of practical problems that can be created by weak instruments can be seen in Angrist and Krueger's (1991) well-known returns to education analysis. Angrist and Krueger investigate returns to schooling in the US using census data from 1960, 1970 and 1980. The authors set out to estimate the effect of an extra year of schooling on subsequent earnings, but this cannot be done with OLS due to the bias potentially introduced by an omitted variable – innate ability – that is thought to positively affect both years spent in education and the earnings. So instead of OLS, Angrist and Krueger used the individual's quarter of birth to instrument out for exogenous variation in schooling which cannot be attributed to the variation in individual intelligence. Due to contemporary legislation that varied by state, students were not allowed to leave school until they reach certain age (16, 17 or 18 years), so some of the variation in the time spent at school was truly exogenous and only explained by the quarter of birth, which was assumed to be uncorrelated with the student's ability. Angrist and Krueger had over 300 000 observations and they used the quarter of birth, year of birth and state dummies, as well as their interactions – altogether more than 170 dummies - for the IV regression.

The coefficients on the quarter of birth dummies in the first stage regression were found to be jointly significantly different from zero at the 1% level, so it was decided that the quarter of birth was a valid instrument. In the second stage the 2SLS estimates were also significant, but insignificantly different from the significant OLS coefficients, which Angrist and Krueger explained by the downward bias in the OLS due to the measurement error or omitted variables. However, another and arguably more likely explanation would be that it was the IV estimator that was biased (towards the OLS estimator) due to the weak IV problem. Angrist and Krueger (1991) is a very important example of the weak IV problem – something that the authors didn't consider when they wrote it, because weak instruments at that time were deemed to be problematic only in small samples.

Bound, Jaeger and Baker (1995) critically assessed the Angrist and Krueger (1991) paper and suggested that the instrument was indeed weak and also potentially not exogenous with respect to wage, which biased the IV results towards the OLS estimates. They managed to reproduce all the coefficients from the original paper and added the first stage F-statistics to different specification and regarded them as being very low.

They also conducted a Monte Carlo study, where they used randomly generated numbers instead of the actual quarter of birth variable and surprisingly got very similar IV results to the Angrist and Krueger (1991) paper. Even the standard errors were very similar and low – it was a common belief that the standard errors would be large if the instruments were weak. And in general, this is true - as the instrument becomes weaker, the standard errors get larger. However, when the instruments are sufficiently weak, larger standard errors do not compensate sufficiently for the bias of the IV estimates, so the inference is not properly sized; it's not just that the estimates get imprecise, but they are also badly biased. However, Bound et al. noted that the first stage F-statistic for the regressions with the randomly generated instruments were very low, almost equal to 1, which suggests that the instruments are not correlated with schooling.

This can be reviewed in terms of the concentration parameter – Stock et al. (2002) discuss the link between the concentration parameter and the F-statistics from testing the hypothesis that $\Pi = 0$ and they show that when the sample size is large:

$$E(F) \cong \frac{\mu^2}{L} + 1 \tag{2.7}$$

This means that $\frac{\mu^2}{L}$ can be estimated by F-1, which essentially means that the F-stat close to 1 corresponds to the concentration parameter that is close to zero, which as discussed earlier, would cause the pdf of the IV estimator to be centered around the same wrong value as the OLS estimator. In light of their findings, Bound et al (1995) emphasized the importance of reporting the diagnostic tests from the first stage IV regression, such as the F-stat and the partial R-squared of the exogenous instruments.

2.4 Detecting weak instruments.

What is a weak instrument? It is difficult to give a general definition because the answer varies by context. It is always true that the more weakly correlated the instruments are with the regressors, the worse are the asymptotic approximations of the distributions of the IV estimator and other related statistics. But what matters is the magnitude of the distortions that could be invoked. Stock and Yogo (2005) suggested measuring these distortions in terms of the maximal bias of the IV estimator relative to the OLS estimator and the maximal size distortions of the Wald test for testing hypotheses on β that a researcher is willing to tolerate.

For example, if the maximal relative bias (of IV compared to OLS) that a researcher is willing to accept is 15% and a maximal size of the Wald test (when the nominal size is 5%) is 20%, then the instrument that generates a larger bias or distorts the size more should be deemed weak.

We established earlier that the value of the concentration parameter is related to the quality of asymptotic approximation, so the question can be reinterpreted as how small should the value of the concentration parameter be for the researcher to conclude that his instruments are weak?

Stock and Yogo (2005) considered a test based on the Cragg-Donald (1993) test statistic g_{min}, which was originally meant for testing the rank of the Π matrix to rule out underidentification, where $g_{min} = mineval(G_n)$ and G_n is a matrix analog of the first–stage F-statistic for the general case with several endogenous regressors and included insruments. Note, that I will also refer to the Cragg-Donald Wald statistic as the first stage F-statistic from testing the joint significance of the instruments in the reduced form equation as in my model with one endogenous regressor the two are equivalent.

Using the results from Staiger and Stock (1997) Stock and Yogo (2005) show that under weak asymptotics, g_{min} converges in distribution to $mineval(\frac{\nu}{L})$, where ν has non-central Wishart distribution with non-centrality parameter equal to the weak instrument limit of the concentration matrix (matrix version of the concentration parameter for a more general IV model). The minimum eigenvalues of the concentration matrix are next used to describe weak instrument sets and derive the critical values associated with a certain maximal bias and maximal test size distortion.

Stock and Yogo (2005) haven't modified the Cragg-Donald statistic, but derived critical values that correspond to a certain relative bias or the test size distortion. The critical values are different depending on which estimator is used (2SLS, LIML, fuller-k etc.), the number of endogenous regressors and the number of instruments. Stata ivreg2 command (Baum et al., 2007) routinely reports the critical values for the first stage F-statistics, when they exist.

It's worth noting that the critical values were derived for the case when ε and ν are assumed to be iid. It is likely to be the case that the critical values would be different in the presence of heteroskedasticity or serial correlation, even if the robust version of the Cragg-Donald statistic – Kliebergen-Paap rk LM statistic, proposed by Kleibergen and Paap (2006) - is used. Olea and Pflueger (2013) proposed a scaled version of the usual F-statistic to test for weak instruments, which is robust to heteroscedasticity and serial correlation. In the exactly identified case with a single endogenous regressor their test statistic is identical to the robust first stage F-statistic, however, in case of overidentification, the test statistic is different.

Before Stock and Yogo (2005) derived the critical values and suggested their version of a weak IV test, it used to be much more common to simply test for underidentification, i.e. test whether the rank condition holds. Other alternatives included looking at some form of the R-squared from the first stage regression, for example, Hall at el. (1996) suggested to use Anderson's (1951) canonical correlations test in the context of weak IV, which in principle is a test of underidentification.[2] As for the tests that specifically check for the presence of weak instruments - there have been very few. A closely-related (but inferior) test to the one proposed by Stock and Yogo (2005), was proposed by Staiger and Stock (1997). They advised using a rule of thumb when working with the models with one endogenous regressor, which stated that if the first stage F-statistic is larger than 10, then the instruments should not be considered weak. Another test, proposed by Hahn and Hausman (2002), has an null hypothesis of strong instruments,

[2] Shea's (1997) partial R-squared can be used in the case of multiple endogenous regressors

so rejecting it would suggest that instruments are weak or irrelevant, but the test had very low power, so it hasn't been popular in applied work.

2.5 Potential problem with the pre-test.

In the previous section I described the test suggested by Stock and Yogo (2005) for detecting weak instruments based on the critical values for the first stage F-statistic. The paper has been very influential and has received 3148 citations[3], which indicates that the method is widely used.

How does the test work? The basic idea is that the researcher decides on the maximal size distortion he is willing to accept – for instance: 15% – obtains the first stage F-stat, compares it to the corresponding critical value – for the case with one endogenous regressor and one instrument it would be 8.96 – and if the F-statistic exceeds the critical value, then the instrument is regarded as not being weak and the researcher can proceed with the IV estimation of the main equation. However, if the F-statistic is below the critical value, the instrument is considered to be weak, so the researcher is better off finding other instruments or undertaking some form of the weak IV robust inference, such as using an Anderson and Rubin (AR) (1949) approach, or a more powerful Kleibergen's (2002) Lagrange Multiplier statistic, or Moreira's (2003) conditional likelihood ratio tests, the distributions of which do not depend on the value of the concentration parameter. There has been extensive research on various modifications of these tests, however, this is not the focus of this paper, an accessible overview can be found in Mikusheva (2013).

Both Staiger and Stock (1997) and Stock and Yogo (2005) tests seem easy to implement and very informative, so what can go wrong?

Hall et al. (1996) suggested that using a pre-test can be detrimental for the IV estimation and can compromise the results even more than a weak IV problem. They consider a likelihood ratio test based on the canonical correlations between the endogenous regressor and the instruments. In the case with one endogenous regressor and one instrument the test statistic is $\Phi = -\mathrm{n} * \log(1 - r^2)$, where r^2 is a square of the

[3] According to Google Scholar, as of 30.07.16

correlation between the instrument and the regressor. Under the null hypothesis of the instruments being uncorrelated with the endogenous regressor Φ is distributed as chi-square with one degree of freedom. Hall et al. (1996) test is closely related to the nR^2 test proposed by Nelson and Startz (1990a), where they suggest that nR^2 larger than 2 corresponds to strong instruments. Hall et al. (1996) point out that when the correlation is low, Φ and nR^2 are approximately equal to each other.

Hall et al (1996) consider the proposed test in the context of a hypothetical situation in which the researcher has access to lists of potential instruments and wants to choose a list that would produce the most reliable inference in the second stage of the IV regression based on the results of the canonical correlation test. He obtains the R-squared from the first stage regression, calculates test-statistic, compares it to the appropriate critical value and if the test statistic doesn't exceed the critical value, the instruments are thrown away. The researcher continues testing the instruments until he can reject the null hypothesis of irrelevant instruments and then those instruments are kept to estimate the structural equation.

They conducted a Monte Carlo study for a simple case with one endogenous regressor and one instrument using 100 observations and 10 000 simulations. The data generated varied in the degree of correlation between the instrument and the regressor from 0 to 0.4 and the degree of endogeneity from 0 to 0.9. One of the points of the experiment was to check how well their Φ-statistic performs as a pre-estimation screening procedure, so they calculated the size for the test of the hypothesis $\beta = 0$ for all the samples, and then for the samples where Φ exceeded the 10% and 1% critical value. The results were worrying – it looked like the samples with a higher Φ-statistic and not very strong correlation between the instrument and the regressor consistently had worse distortions to the test size, and the distortions increased steadily as the degree of endogeneity was getting stronger. Their conclusion was that the researcher seems to be better off using a random instrument rather than an instrument that passes the pre-test. The explanation they provided is straightforward – by using a pre-test, the researcher introduces a pre-test bias into the IV estimation.

In the next section I describe an experiment I conducted based on the Hall et al. (1996) design and Stock and Yogo's critical values to test whether Stock and Yogo's (2005) weak instrument test and critical values suffer from the pre-test bias in a similar way to the Hall's Φ-statistic and test for underidentification. I expect to find analogous results due to the fact that it can be easily shown that for the case with one endogenous regressor for the fixed n and fixed l there is a one-to-one relationship between the F-statistic and the R^2 from the first stage regression:

$$F = \frac{R^2/(l-1)}{(1-R^2)/(n-l)} \tag{2.8}$$

2.6 Design of the Monte Carlo simulations and the content of the experiment.

For the experiment I will use the model described earlier in section 2.2,

$$y_i = x_i\beta + \varepsilon_i, \tag{2.9}$$

$$x_i = \boldsymbol{z}_i\pi + v_i, \tag{2.10}$$

All variables are randomly drawn from the normal distribution with zero mean and variance-covariance matrix V, true β is set to 0. For all the simulations I use 10 000 repetitions.

$$V = \begin{pmatrix} {\sigma^2}_x & \sigma_{xz} & \sigma_{x\varepsilon} \\ - & {\sigma^2}_z & \sigma_{z\varepsilon} \\ - & - & {\sigma^2}_\varepsilon \end{pmatrix}$$

${\sigma^2}_x$=${\sigma^2}_z$=${\sigma^2}_\varepsilon$=1, $\sigma_{z\varepsilon}$=0

I will consider three cases:

Case 1: 1 endogenous regressor, 1 excluded instrument, $n = 100$, σ_{xz}varies from 0 to 0.4, $\sigma_{x\varepsilon}$ varies from 0 to 0.9

Case 2: 1 endogenous regressor, 2 excluded instruments, $n = 100$, ${\sigma^2}_{z_1} = {\sigma^2}_{z_2}$, $\sigma_{z_1\varepsilon} = \sigma_{z_2\varepsilon} = 0$, $\sigma_{xz_1} = \sigma_{xz_2} = \sigma_{xz}$ and vary from 0 to 0.4, $\sigma_{x\varepsilon}$ varies from 0 to 0.9

Case 3: 1 endogenous regressor, 1 excluded instrument, $n = 1600$, σ_{xz}varies from 0 to 0.1, $\sigma_{x\varepsilon}$ varies from 0 to 0.9

In each round of the simulations the reduced form equation is estimated first, the F-statistics is collected, then the structural equation is estimated, $H_0: \beta = 0$ tested by means of Wald test with the 5% nominal size and the p-value recorded.

The main purpose of the experiment is to test whether the samples with the first stage F-statistic above Stock and Yogo's critical values will produce lower Wald test size distortions, specifically if the actual size in the finite samples will match the "predicted" size, when compared to randomly selected samples.

The aggregated results are presented in tables 2.1, 2.2 and 2.3 that correspond to the cases described above. I have also calculated the median-squared error, which is used as a loss function, for all the samples and the samples conditional on the first stage F-statistic exceeding the critical value. The median-squared error is used instead of more traditional mean-squared error since in the exactly identified case the IV estimator does not have any finite moments - the moments of the IV estimator only exist up to the degree of overidentification (see Mariano (1972)).[4]

All the tables are structured in the same way:

Column (1) defines the strength of the correlation between the endogenous regressor and the instrument, column (2) show the degree of endogeneity of x, columns (3)-(5) report the values for the estimated 5%, 50% and 95% fractiles for the obtained distribution of $\hat{\beta}$, column (6) reports size of the Wald test for all samples, column (7) – median F-statistic for all samples, columns (8)-(17) report the fractions of all samples that exceed the conventional 5% critical value, the F-statistic above 10 (Staiger and Stock's rule of thumb) and then F-statistics of 5.53, 8.96 and 16.38, which are the Stock and Yogo's critical values for 25%, 15% and 20% maximal Wald test size and the actual test sizes observed in these sample (the critical values are different for the case with two instruments). Columns (18)-(20) report the median-squared error for all the samples and for samples conditional on passing Stock and Yogo's weak IV test. Each table contains five panels (A-E) and each panel corresponds to different strength of the relationship between x and z, from completely unrelated to strongly correlated.

[4] However, similar results were obtained using the mean-squared error as a loss function

2.7 Discussion of the results (2SLS).

Panel A and panel B of table 2.1 correspond to no correlation between x and z and very low correlation with σ_{xz}=0.1, which is supported by very low median F-statistics for all samples – the maximum value it reaches is 1.14, which is far below the 5% critical value of 3.94. Panels C, D and E produce median F-statistics with values just above the 5% critical values when $\sigma_{xz} = 0.2$ and the F-stat of almost 20 when σ_{xz}=0.4.

Another thing to notice is that within panels A and B as the correlation between x and ε goes up, the median of the observed distribution of β also goes up, so it is clear that the 2SLS estimator is biased towards the OLS and when the instrument is irrelevant or weak the 2SLS bias gets closer to the OLS bias as the degree of endogeneity increases. We can still observe this pattern in panel C, however the bias is notably smaller and for the samples with $\sigma_{xz} \geq 0.3$ the bias basically disappears for any degree of endogeneity. Also, as the strength of the instrument increases, the observed distributions of $\hat{\beta}$ get narrower around the true value of β. The distributions also seem fairly symmetric in panels A and B, but become left-skewed in Panles C, D and E, even despite the fact that for those samples the median of the distribution is close to the true value of zero. This suggests that there were more extreme negative than positive $\hat{\beta}$s estimated.

Table 2.1. Case 1: 2SLS, n=100, 1 endogenous regressor, 1 instrument.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$					Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95	Size of stat($\hat{\beta}$) (all samples)	Median F-stat	Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	All samples	Fstat>8.96	Fstat>16.38
Panel A	0.0	0.0	-6.411	-0.004	5.750	0.000	0.458	0.053	0.004	0.002	0.000	0.022	0.009	0.004	0.000	0.000	0.000	0.990	0.058	
	0.0	0.1	-6.277	0.112	6.363	0.000	0.458	0.050	0.002	0.002	0.043	0.019	0.005	0.003	0.031	0.000	0.000	0.998	0.032	
	0.0	0.2	-6.337	0.199	5.963	0.001	0.463	0.050	0.016	0.002	0.105	0.021	0.033	0.003	0.065	0.000	0.333	0.985	0.044	
	0.0	0.3	-5.593	0.305	6.509	0.002	0.465	0.052	0.035	0.002	0.143	0.020	0.070	0.003	0.207	0.000	0.000	0.998	0.148	
	0.0	0.4	-5.372	0.410	5.995	0.005	0.458	0.049	0.084	0.002	0.278	0.018	0.133	0.003	0.206	0.000	0.000	0.986	0.234	
	0.0	0.5	-5.364	0.500	5.735	0.016	0.476	0.050	0.223	0.001	0.250	0.020	0.288	0.002	0.304	0.000	0.333	0.964	0.198	0.094
	0.0	0.6	-4.489	0.599	5.681	0.031	0.457	0.051	0.371	0.002	0.650	0.021	0.483	0.004	0.657	0.000	1.000	1.002	0.376	0.115
	0.0	0.7	-3.520	0.718	5.448	0.066	0.461	0.053	0.561	0.002	0.765	0.022	0.653	0.003	0.720	0.000	1.000	1.022	0.413	0.316
	0.0	0.8	-2.803	0.789	4.326	0.129	0.464	0.049	0.816	0.002	0.889	0.019	0.880	0.004	0.947	0.000	0.000	1.014	0.631	0.944
	0.0	0.9	-1.740	0.914	3.551	0.262	0.459	0.049	0.969	0.002	1.000	0.019	0.989	0.004	1.000	0.000	0.000	0.992	0.785	0.695

Table 2.1. Case 1: 2SLS, n=100, 1 endogenous regressor, 1 instrument (continued).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$			Size of stat($\hat{\beta}$) (all samples)	Median F-stat	Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95			Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	All samples	Fstat> 8.96	Fstat> 16.38
Panel B	0.1	0.0	-3.922	-0.001	3.871	0.001	1.103	0.173	0.003	0.018	0.022	0.094	0.006	0.027	0.015	0.001	0.000	0.424	0.040	0.018
	0.1	0.1	-3.919	0.022	4.021	0.002	1.116	0.172	0.009	0.019	0.046	0.094	0.016	0.028	0.043	0.002	0.043	0.433	0.039	0.039
	0.1	0.2	-3.750	0.084	3.957	0.003	1.089	0.171	0.018	0.019	0.063	0.097	0.025	0.026	0.060	0.002	0.048	0.434	0.059	0.015
	0.1	0.3	-4.277	0.119	4.025	0.006	1.109	0.163	0.036	0.015	0.125	0.086	0.054	0.024	0.106	0.002	0.200	0.424	0.069	0.024
	0.1	0.4	-3.778	0.150	3.723	0.013	1.142	0.169	0.069	0.017	0.178	0.096	0.094	0.025	0.165	0.002	0.190	0.418	0.086	0.198
	0.1	0.5	-3.757	0.222	4.107	0.028	1.106	0.171	0.144	0.018	0.249	0.095	0.178	0.027	0.259	0.002	0.381	0.444	0.110	0.109
	0.1	0.6	-3.679	0.229	4.029	0.040	1.118	0.171	0.194	0.015	0.379	0.094	0.240	0.026	0.320	0.001	0.667	0.419	0.147	0.227
	0.1	0.7	-3.693	0.273	4.132	0.076	1.133	0.174	0.319	0.020	0.594	0.101	0.406	0.030	0.547	0.002	0.688	0.390	0.218	0.353
	0.1	0.8	-3.808	0.298	4.259	0.112	1.143	0.170	0.410	0.018	0.737	0.094	0.511	0.025	0.723	0.002	0.950	0.359	0.317	0.320
	0.1	0.9	-4.394	0.291	4.712	0.163	1.109	0.167	0.617	0.017	0.971	0.094	0.774	0.025	0.952	0.002	1.000	0.328	0.397	0.437
Panel C	0.2	0.0	-1.312	0.007	1.356	0.004	4.123	0.516	0.008	0.139	0.018	0.375	0.011	0.176	0.016	0.030	0.010	0.124	0.043	0.019
	0.2	0.1	-1.398	0.008	1.205	0.006	4.097	0.515	0.011	0.137	0.024	0.376	0.015	0.176	0.023	0.029	0.055	0.121	0.039	0.033
	0.2	0.2	-1.459	0.003	1.105	0.010	4.124	0.516	0.020	0.134	0.047	0.378	0.026	0.170	0.042	0.026	0.080	0.125	0.039	0.028
	0.2	0.3	-1.551	0.025	1.045	0.017	4.215	0.523	0.032	0.136	0.065	0.383	0.039	0.176	0.056	0.029	0.103	0.122	0.042	0.043
	0.2	0.4	-1.635	0.025	0.948	0.025	4.213	0.528	0.048	0.138	0.101	0.388	0.060	0.177	0.093	0.027	0.179	0.114	0.045	0.038
	0.2	0.5	-1.751	0.016	0.839	0.041	4.175	0.522	0.075	0.140	0.155	0.383	0.096	0.182	0.138	0.026	0.277	0.119	0.059	0.075
	0.2	0.6	-2.015	0.017	0.743	0.055	4.185	0.524	0.101	0.141	0.216	0.386	0.124	0.177	0.198	0.029	0.358	0.114	0.073	0.099
	0.2	0.7	-2.132	0.032	0.680	0.073	4.248	0.531	0.134	0.143	0.288	0.387	0.166	0.182	0.257	0.029	0.502	0.112	0.083	0.131
	0.2	0.8	-2.275	0.027	0.596	0.088	4.134	0.519	0.163	0.136	0.400	0.379	0.211	0.174	0.358	0.026	0.629	0.109	0.111	0.172
	0.2	0.9	-2.592	0.025	0.536	0.096	4.155	0.522	0.182	0.137	0.525	0.376	0.248	0.175	0.453	0.027	0.882	0.106	0.141	0.227

Table 2.1. Case 1: 2SLS, n=100, 1 endogenous regressor, 1 instrument (continued).

	(1) σ_{xz}	(2) σ_{xu}	(3) Estimated fractiles of $\hat{\beta}$: 0.05	(4) Estimated fractiles of $\hat{\beta}$: 0.50	(5) Estimated fractiles of $\hat{\beta}$: 0.95	(6) Size of stat($\hat{\beta}$) (all samples)	(7) Median F-stat	(8) Fstat significant at 5% level: Fraction of all samples	(9) Fstat significant at 5% level: Size of tstat($\hat{\beta}$)	(10) Fstat>10: Fraction of all samples	(11) Fstat>10: Size of tstat($\hat{\beta}$)	(12) Fstat>5.53: Fraction of all samples	(13) Fstat>5.53: Size of tstat($\hat{\beta}$)	(14) Fstat>8.96: Fraction of all samples	(15) Fstat>8.96: Size of tstat($\hat{\beta}$)	(16) Fstat>16.38: Fraction of all samples	(17) Fstat>16.38: Size of tstat($\hat{\beta}$)	(18) Median-squared error, for the samples with: All samples	(19) Median-squared error, for the samples with: Fstat> 8.96	(20) Median-squared error, for the samples with: Fstat> 16.38
Panel D	0.3	0.0	-0.653	0.001	0.652	0.015	9.867	0.869	0.017	0.493	0.027	0.774	0.020	0.555	0.025	0.198	0.035	0.051	0.032	0.022
	0.3	0.1	-0.702	-0.000	0.598	0.015	9.794	0.867	0.018	0.488	0.026	0.776	0.020	0.552	0.024	0.195	0.037	0.052	0.033	0.026
	0.3	0.2	-0.734	0.001	0.560	0.018	9.937	0.873	0.021	0.496	0.029	0.781	0.023	0.560	0.027	0.198	0.039	0.051	0.033	0.024
	0.3	0.3	-0.774	-0.004	0.546	0.024	9.837	0.868	0.028	0.490	0.041	0.775	0.031	0.551	0.039	0.191	0.054	0.053	0.035	0.027
	0.3	0.4	-0.821	-0.002	0.501	0.035	9.843	0.868	0.040	0.490	0.062	0.779	0.044	0.557	0.057	0.202	0.094	0.053	0.035	0.027
	0.3	0.5	-0.910	-0.006	0.465	0.039	9.903	0.867	0.045	0.493	0.071	0.775	0.050	0.559	0.065	0.199	0.116	0.051	0.034	0.035
	0.3	0.6	-0.923	0.002	0.447	0.056	9.908	0.872	0.064	0.495	0.102	0.778	0.071	0.562	0.094	0.198	0.172	0.048	0.033	0.038
	0.3	0.7	-1.000	0.001	0.423	0.065	9.985	0.867	0.075	0.499	0.124	0.780	0.083	0.564	0.112	0.205	0.225	0.050	0.033	0.053
	0.3	0.8	-1.033	0.001	0.396	0.075	10.020	0.872	0.086	0.501	0.147	0.778	0.097	0.561	0.133	0.203	0.299	0.049	0.031	0.060
	0.3	0.9	-1.140	-0.006	0.372	0.079	9.803	0.867	0.092	0.488	0.162	0.773	0.103	0.554	0.143	0.196	0.374	0.048	0.032	0.083
Panel E	0.4	0.0	-0.455	-0.003	0.440	0.023	19.152	0.989	0.023	0.875	0.025	0.971	0.023	0.907	0.025	0.616	0.029	0.029	0.026	0.021
	0.4	0.1	-0.466	0.003	0.419	0.027	19.197	0.989	0.027	0.867	0.030	0.970	0.028	0.898	0.029	0.617	0.034	0.028	0.026	0.022
	0.4	0.2	-0.491	0.004	0.415	0.030	19.044	0.986	0.031	0.869	0.034	0.970	0.031	0.900	0.033	0.615	0.040	0.029	0.026	0.022
	0.4	0.3	-0.511	0.009	0.400	0.033	19.115	0.987	0.033	0.870	0.037	0.970	0.034	0.900	0.036	0.615	0.046	0.028	0.025	0.021
	0.4	0.4	-0.524	0.003	0.377	0.038	19.025	0.987	0.038	0.875	0.043	0.969	0.039	0.903	0.041	0.616	0.054	0.028	0.025	0.020
	0.4	0.5	-0.536	0.002	0.361	0.045	19.102	0.989	0.046	0.871	0.052	0.971	0.047	0.902	0.050	0.618	0.068	0.028	0.024	0.021
	0.4	0.6	-0.579	0.001	0.348	0.055	18.985	0.989	0.056	0.866	0.064	0.973	0.057	0.897	0.061	0.613	0.087	0.027	0.024	0.021
	0.4	0.7	-0.604	0.003	0.332	0.060	18.915	0.986	0.061	0.866	0.069	0.969	0.062	0.896	0.067	0.615	0.097	0.028	0.023	0.020
	0.4	0.8	-0.605	0.002	0.311	0.063	19.129	0.990	0.064	0.877	0.072	0.975	0.065	0.904	0.070	0.625	0.100	0.029	0.023	0.021
	0.4	0.9	-0.657	-0.003	0.299	0.066	18.812	0.989	0.067	0.869	0.076	0.971	0.068	0.902	0.073	0.609	0.109	0.027	0.021	0.017

Table 2.2. Case 2: 2SLS, n=100, 1 endogenous regressor, 2 instruments.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$					Fstat significant at 5% level		Fstat>10		Fstat>7.25		Fstat>11.59		Fstat>19.93		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95	Size of stat(β') (all samples)	Median F-stat Size of tstat(β)	Fraction of all samples	Size of tstat(β')	Fraction of all samples	Size of tstat(β')	Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat(β')	Fraction of all samples	Size of tstat(β')	Fraction of all samples	Fstat> 11.59	Fstat> 19.93
Panel A	0.0	0	-2.123	-0.001	2.010	0.001	0.715	0.055	0.011	0.000	0.000	0.002	0.000	0.000		0.000		0.337		
	0.0	0.1	-1.981	0.096	2.119	0.001	0.711	0.047	0.013	0.000		0.001	0.111	0.000		0.000		0.340		
	0.0	0.2	-1.824	0.213	2.257	0.004	0.693	0.052	0.056	0.000	0.000	0.001	0.154	0.000		0.000		0.356		
	0.0	0.3	-1.711	0.301	2.252	0.007	0.694	0.045	0.099	0.000		0.001	0.375	0.000		0.000		0.380		
	0.0	0.4	-1.473	0.413	2.258	0.019	0.705	0.048	0.203	0.000		0.002	0.250	0.000		0.000		0.418		
	0.0	0.5	-1.287	0.491	2.330	0.045	0.686	0.047	0.406	0.000	1.000	0.001	0.750	0.000		0.000		0.453		
	0.0	0.6	-1.071	0.599	2.319	0.086	0.678	0.048	0.554	0.000	1.000	0.001	0.750	0.000		0.000		0.516		
	0.0	0.7	-0.752	0.698	2.221	0.160	0.696	0.048	0.723	0.000		0.001	1.000	0.000		0.000		0.597		
	0.0	0.8	-0.479	0.798	2.039	0.298	0.705	0.047	0.934	0.000	1.000	0.001	1.000	0.000		0.000		0.702		
	0.0	0.9	0.003	0.903	1.815	0.514	0.710	0.049	1.000	0.000	1.000	0.001	1.000	0.000		0.000		0.837		

Table 2.2. Case 2: 2SLS, n=100, 1 endogenous regressor, 2 instruments (continued).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$					Fstat significant at 5% level		Fstat>10		Fstat>7.25		Fstat>11.59		Fstat>19.93		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95	Size of stat(β′) (all samples)	Median F-stat	Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat(β′)	Fraction of all samples	Size of tstat(β′)	Fraction of all samples	Size of tstat(β′)	Fraction of all samples	Size of tstat(β′)	All samples	Fstat> 11.59	Fstat> 19.93
Panel B	0.1	0	-1.312	-0.005	1.273	0.003	1.597	0.229	0.011	0.003	0.000	0.020	0.034	0.001	0.000	0.000		0.157	0.024	
	0.1	0.1	-1.320	0.047	1.354	0.004	1.547	0.222	0.015	0.003	0.061	0.018	0.040	0.001	0.091	0.000		0.164	0.025	
	0.1	0.2	-1.330	0.086	1.328	0.009	1.555	0.222	0.037	0.004	0.050	0.020	0.085	0.002	0.105	0.000		0.167	0.047	
	0.1	0.3	-1.159	0.144	1.383	0.018	1.567	0.225	0.067	0.004	0.143	0.020	0.127	0.001	0.077	0.000		0.165	0.030	
	0.1	0.4	-1.195	0.186	1.406	0.034	1.570	0.224	0.122	0.003	0.312	0.017	0.224	0.002	0.333	0.000		0.175	0.096	
	0.1	0.5	-1.105	0.228	1.433	0.061	1.606	0.225	0.182	0.003	0.424	0.020	0.332	0.001	0.455	0.000		0.177	0.125	
	0.1	0.6	-1.066	0.266	1.344	0.095	1.577	0.225	0.261	0.004	0.703	0.019	0.511	0.001	0.800	0.000		0.182	0.213	
	0.1	0.7	-1.001	0.308	1.369	0.147	1.568	0.223	0.369	0.003	0.926	0.019	0.718	0.001	1.000	0.000	1.000	0.193	0.232	0.199
	0.1	0.8	-0.949	0.350	1.374	0.213	1.575	0.221	0.522	0.004	0.946	0.018	0.844	0.001	1.000	0.000		0.207	0.408	
	0.1	0.9	-0.783	0.380	1.411	0.291	1.599	0.224	0.688	0.004	1.000	0.019	0.979	0.002	1.000	0.000		0.205	0.417	
Panel C	0.2	0	-0.643	-0.005	0.642	0.012	4.874	0.736	0.017	0.094	0.028	0.252	0.025	0.054	0.035	0.002	0.105	0.053	0.021	0.006
	0.2	0.1	-0.658	0.013	0.637	0.015	4.872	0.727	0.021	0.102	0.031	0.254	0.032	0.055	0.027	0.001	0.071	0.054	0.020	0.006
	0.2	0.2	-0.681	0.030	0.614	0.023	4.905	0.740	0.031	0.102	0.062	0.254	0.052	0.056	0.072	0.003	0.133	0.052	0.026	0.019
	0.2	0.3	-0.673	0.040	0.594	0.035	4.855	0.739	0.046	0.105	0.094	0.255	0.080	0.058	0.119	0.002	0.062	0.056	0.031	0.009
	0.2	0.4	-0.677	0.052	0.562	0.046	4.885	0.737	0.061	0.100	0.143	0.256	0.105	0.055	0.161	0.002	0.158	0.051	0.039	0.027
	0.2	0.5	-0.708	0.062	0.554	0.060	4.782	0.729	0.078	0.098	0.196	0.250	0.141	0.056	0.219	0.002	0.417	0.055	0.053	0.072
	0.2	0.6	-0.744	0.076	0.529	0.082	4.895	0.737	0.108	0.106	0.293	0.259	0.201	0.057	0.351	0.002	0.812	0.055	0.076	0.136
	0.2	0.7	-0.723	0.090	0.499	0.101	4.908	0.741	0.132	0.095	0.405	0.250	0.279	0.054	0.489	0.001	0.786	0.054	0.100	0.118
	0.2	0.8	-0.711	0.104	0.488	0.121	4.866	0.733	0.160	0.094	0.535	0.248	0.345	0.051	0.619	0.002	0.944	0.054	0.123	0.223
	0.2	0.9	-0.741	0.116	0.452	0.137	4.831	0.736	0.183	0.096	0.752	0.252	0.456	0.055	0.868	0.002	1.000	0.053	0.168	0.219

Table 2.2. Case 2: 2SLS, n=100, 1 endogenous regressor, 2 instruments (continued).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$			Size of stat($\hat{\beta}$) (all samples)	Median F-stat	Fstat significant at 5% level		Fstat>10		Fstat>7.25		Fstat>11.59		Fstat>19.93		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95			Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	All samples	Fstat> 11.59	Fstat> 19.93
Panel D	0.3	0	-0.398	0.000	0.406	0.030	11.594	0.987	0.030	0.627	0.034	0.826	0.033	0.500	0.035	0.092	0.045	0.025	0.018	0.013
	0.3	0.1	-0.418	0.002	0.395	0.032	11.610	0.987	0.032	0.621	0.036	0.820	0.035	0.501	0.039	0.090	0.048	0.024	0.018	0.013
	0.3	0.2	-0.407	0.011	0.383	0.032	11.531	0.986	0.032	0.622	0.040	0.822	0.035	0.495	0.045	0.096	0.063	0.024	0.018	0.014
	0.3	0.3	-0.433	0.015	0.378	0.040	11.549	0.987	0.040	0.625	0.053	0.826	0.046	0.497	0.058	0.093	0.088	0.025	0.019	0.018
	0.3	0.4	-0.438	0.018	0.365	0.045	11.550	0.987	0.046	0.618	0.064	0.818	0.053	0.497	0.070	0.089	0.126	0.025	0.019	0.021
	0.3	0.5	-0.442	0.028	0.360	0.053	11.457	0.986	0.054	0.610	0.080	0.819	0.063	0.488	0.093	0.088	0.205	0.024	0.020	0.029
	0.3	0.6	-0.474	0.031	0.348	0.066	11.566	0.985	0.067	0.620	0.101	0.820	0.080	0.498	0.118	0.092	0.276	0.025	0.022	0.039
	0.3	0.7	-0.466	0.035	0.335	0.072	11.521	0.986	0.073	0.610	0.115	0.819	0.087	0.495	0.137	0.088	0.358	0.024	0.023	0.050
	0.3	0.8	-0.471	0.049	0.337	0.090	11.502	0.988	0.092	0.614	0.145	0.815	0.111	0.494	0.178	0.093	0.492	0.026	0.028	0.067
	0.3	0.9	-0.486	0.050	0.317	0.095	11.467	0.985	0.097	0.611	0.156	0.813	0.117	0.492	0.193	0.091	0.648	0.025	0.031	0.081
Panel E	0.4	0	-0.298	0.000	0.298	0.041	23.878	1.000	0.041	0.985	0.042	0.997	0.041	0.966	0.042	0.690	0.041	0.014	0.014	0.012
	0.4	0.1	-0.301	0.001	0.290	0.040	24.133	1.000	0.040	0.984	0.040	0.997	0.040	0.968	0.040	0.700	0.043	0.014	0.014	0.012
	0.4	0.2	-0.301	0.007	0.291	0.043	24.065	1.000	0.043	0.987	0.043	0.997	0.043	0.971	0.043	0.702	0.047	0.014	0.014	0.012
	0.4	0.3	-0.316	0.012	0.282	0.046	24.152	1.000	0.046	0.984	0.046	0.996	0.046	0.967	0.046	0.705	0.052	0.014	0.013	0.012
	0.4	0.4	-0.317	0.009	0.273	0.047	23.870	1.000	0.047	0.984	0.047	0.997	0.047	0.965	0.047	0.697	0.059	0.014	0.014	0.012
	0.4	0.5	-0.324	0.016	0.269	0.051	24.036	1.000	0.051	0.984	0.051	0.997	0.051	0.966	0.052	0.703	0.068	0.014	0.014	0.012
	0.4	0.6	-0.335	0.016	0.265	0.056	23.874	1.000	0.056	0.982	0.057	0.998	0.056	0.963	0.058	0.693	0.078	0.014	0.013	0.012
	0.4	0.7	-0.322	0.023	0.260	0.064	24.075	1.000	0.064	0.985	0.064	0.997	0.064	0.968	0.065	0.702	0.090	0.014	0.014	0.012
	0.4	0.8	-0.341	0.023	0.255	0.069	24.107	1.000	0.069	0.985	0.070	0.998	0.069	0.969	0.071	0.699	0.098	0.014	0.013	0.011

Table 2.3. Case 3: 2SLS, n=1600, 1 endogenous regressor, 1 instrument.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$			Size of stat($\hat{\beta}$) (all samples)	Median F-stat Size of tstat(β)	Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95			Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Fstat> 8.96	Fstat> 16.38
Panel A	0.0	0.0	-5.832	-0.009	6.745	0.000	0.447	0.048	0.002	0.019	0.005	0.001	0.083	0.002	0.050	0.000	0.000	1.000	0.069	0.139
	0.0	0.1	-6.278	0.113	6.777	0.000	0.446	0.054	0.007	0.018	0.022	0.002	0.059	0.003	0.094	0.000	1.000	0.997	0.084	0.235
	0.0	0.2	-5.725	0.187	6.364	0.001	0.455	0.049	0.012	0.021	0.029	0.001	0.091	0.002	0.045	0.000		0.943	0.035	
	0.0	0.3	-5.732	0.308	6.741	0.003	0.467	0.046	0.065	0.016	0.110	0.001	0.214	0.003	0.179	0.000	0.000	1.046	0.138	0.107
	0.0	0.4	-5.954	0.399	6.187	0.008	0.444	0.049	0.147	0.019	0.245	0.002	0.381	0.003	0.313	0.000	0.000	0.957	0.139	0.006
	0.0	0.5	-5.569	0.483	5.933	0.014	0.438	0.050	0.218	0.019	0.301	0.002	0.412	0.003	0.400	0.000		1.038	0.194	
	0.0	0.6	-4.306	0.592	5.385	0.030	0.467	0.049	0.382	0.017	0.434	0.001	0.667	0.002	0.529	0.000		0.987	0.261	
	0.0	0.7	-3.886	0.702	5.444	0.064	0.464	0.049	0.600	0.019	0.661	0.001	0.769	0.002	0.652	0.000		1.008	0.396	
	0.0	0.8	-2.934	0.799	4.443	0.135	0.472	0.052	0.813	0.021	0.917	0.002	1.000	0.003	0.969	0.000		0.992	0.670	
	0.0	0.9	-1.907	0.897	3.689	0.265	0.458	0.052	0.981	0.020	1.000	0.001	1.000	0.003	1.000	0.000		0.992	0.812	

Table 2.3. Case 3: 2SLS, n=1600, 1 endogenous regressor, 1 instrument (continued).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$					Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95	Size of stat(β') (all samples)	Median F-stat	Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat(β')	Fraction of all samples	Size of tstat(β')	Fraction of all samples	Size of tstat(β')	Fraction of all samples	Size of tstat(β')	All samples	Fstat> 8.96	Fstat> 16.38
Panel B	0.025	0.0	-3.888	0.011	3.988	0.001	1.063	0.172	0.004	0.092	0.008	0.015	0.026	0.024	0.025	0.001	0.000	0.459	0.046	0.007
	0.025	0.1	-4.154	0.045	3.741	0.001	1.098	0.166	0.006	0.090	0.009	0.017	0.035	0.026	0.026	0.001	0.000	0.452	0.034	0.016
	0.025	0.2	-3.848	0.081	3.883	0.002	1.138	0.169	0.013	0.090	0.022	0.016	0.051	0.024	0.054	0.001	0.100	0.439	0.040	0.050
	0.025	0.3	-3.890	0.112	4.075	0.005	1.082	0.172	0.031	0.090	0.051	0.017	0.106	0.025	0.081	0.001	0.167	0.434	0.068	0.105
	0.025	0.4	-3.812	0.171	4.103	0.011	1.082	0.168	0.066	0.085	0.107	0.015	0.141	0.024	0.127	0.001	0.182	0.445	0.086	0.085
	0.025	0.5	-3.940	0.192	3.718	0.022	1.099	0.167	0.122	0.088	0.164	0.015	0.240	0.023	0.228	0.001	0.200	0.415	0.117	0.095
	0.025	0.6	-3.762	0.247	4.233	0.039	1.102	0.169	0.192	0.088	0.271	0.014	0.403	0.022	0.403	0.001	0.500	0.408	0.155	0.140
	0.025	0.7	-3.963	0.271	4.142	0.067	1.082	0.167	0.288	0.085	0.394	0.015	0.626	0.021	0.561	0.001	0.636	0.375	0.233	0.254
	0.025	0.8	-4.010	0.289	4.093	0.112	1.093	0.173	0.421	0.090	0.525	0.016	0.831	0.023	0.794	0.002	1.000	0.368	0.327	0.381
	0.025	0.9	-3.807	0.298	4.678	0.153	1.105	0.165	0.595	0.086	0.756	0.017	0.959	0.023	0.939	0.001	1.000	0.327	0.410	0.487
Panel C	0.05	0.0	-1.262	0.001	1.303	0.004	4.068	0.523	0.007	0.370	0.009	0.124	0.019	0.166	0.016	0.021	0.034	0.131	0.039	0.025
	0.05	0.1	-1.335	0.016	1.189	0.004	4.081	0.525	0.008	0.367	0.012	0.125	0.024	0.166	0.022	0.022	0.032	0.123	0.040	0.031
	0.05	0.2	-1.482	0.017	1.102	0.009	4.078	0.523	0.016	0.371	0.022	0.124	0.042	0.163	0.037	0.021	0.038	0.127	0.038	0.025
	0.05	0.3	-1.541	0.008	0.989	0.016	4.005	0.516	0.031	0.363	0.042	0.125	0.072	0.158	0.067	0.021	0.121	0.124	0.042	0.051
	0.05	0.4	-1.708	0.021	0.956	0.025	4.044	0.518	0.048	0.367	0.064	0.122	0.102	0.159	0.092	0.018	0.141	0.123	0.050	0.055
	0.05	0.5	-1.814	0.019	0.876	0.036	3.919	0.507	0.070	0.356	0.093	0.127	0.153	0.163	0.132	0.020	0.255	0.117	0.055	0.070
	0.05	0.6	-2.005	0.014	0.758	0.055	4.001	0.517	0.102	0.360	0.133	0.126	0.228	0.161	0.209	0.021	0.405	0.118	0.074	0.113
	0.05	0.7	-2.189	0.022	0.677	0.068	3.980	0.513	0.125	0.364	0.162	0.126	0.305	0.163	0.264	0.020	0.562	0.107	0.090	0.145
	0.05	0.8	-2.185	0.020	0.598	0.087	3.993	0.515	0.166	0.362	0.221	0.121	0.435	0.159	0.378	0.021	0.756	0.108	0.118	0.186
	0.05	0.9	-2.672	0.015	0.544	0.100	3.907	0.506	0.195	0.360	0.270	0.120	0.590	0.159	0.511	0.020	0.951	0.105	0.150	0.251

Table 2.3. Case 3: 2SLS, n=1600, 1 endogenous regressor, 1 instrument (continued).

Panel	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$					Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95	Size of stat(β̂') (all samples)	Median F-stat	Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat(β̂')	Fraction of all samples	Size of tstat(β̂')	Fraction of all samples	Size of tstat(β̂')	Fraction of all samples	Size of tstat(β̂')	All samples	Fstat> 8.96	Fstat> 16.38
Panel D	0.075	0.0	-0.661	-0.005	0.651	0.012	9.059	0.847	0.014	0.742	0.016	0.443	0.023	0.507	0.021	0.149	0.037	0.054	0.032	0.020
	0.075	0.1	-0.707	-0.005	0.593	0.011	9.033	0.856	0.013	0.746	0.015	0.435	0.023	0.504	0.021	0.150	0.025	0.050	0.030	0.020
	0.075	0.2	-0.754	0.003	0.591	0.018	8.912	0.838	0.021	0.733	0.024	0.433	0.034	0.497	0.031	0.151	0.046	0.052	0.032	0.023
	0.075	0.3	-0.771	0.008	0.552	0.029	9.070	0.855	0.033	0.748	0.037	0.437	0.053	0.509	0.049	0.150	0.078	0.052	0.033	0.026
	0.075	0.4	-0.835	-0.000	0.522	0.034	9.026	0.852	0.039	0.748	0.044	0.437	0.064	0.504	0.059	0.148	0.099	0.053	0.034	0.032
	0.075	0.5	-0.903	0.006	0.487	0.047	8.998	0.851	0.055	0.737	0.063	0.435	0.092	0.503	0.084	0.149	0.150	0.053	0.034	0.035
	0.075	0.6	-0.924	0.003	0.447	0.051	9.174	0.852	0.060	0.747	0.068	0.444	0.102	0.516	0.092	0.147	0.184	0.051	0.033	0.042
	0.075	0.7	-0.999	0.002	0.421	0.063	9.006	0.850	0.074	0.743	0.084	0.434	0.136	0.502	0.120	0.148	0.270	0.047	0.032	0.058
	0.075	0.8	-1.062	-0.003	0.396	0.072	9.135	0.852	0.084	0.745	0.096	0.444	0.155	0.511	0.137	0.152	0.345	0.047	0.032	0.074
	0.075	0.9	-1.122	-0.006	0.373	0.077	8.915	0.850	0.090	0.738	0.104	0.430	0.178	0.498	0.154	0.149	0.434	0.048	0.033	0.089
Panel E	0.1	0.0	-0.454	-0.004	0.459	0.022	16.062	0.981	0.023	0.955	0.023	0.805	0.026	0.852	0.025	0.484	0.029	0.030	0.026	0.021
	0.1	0.1	-0.476	0.005	0.430	0.025	16.206	0.980	0.025	0.954	0.026	0.803	0.029	0.847	0.028	0.492	0.036	0.030	0.026	0.021
	0.1	0.2	-0.496	-0.001	0.415	0.026	16.092	0.978	0.026	0.949	0.027	0.799	0.031	0.843	0.030	0.486	0.038	0.029	0.025	0.020
	0.1	0.3	-0.512	-0.001	0.392	0.031	16.162	0.981	0.031	0.953	0.032	0.806	0.037	0.847	0.036	0.490	0.050	0.028	0.024	0.020
	0.1	0.4	-0.539	0.001	0.378	0.035	16.258	0.980	0.036	0.953	0.037	0.799	0.043	0.844	0.041	0.494	0.059	0.029	0.024	0.021
	0.1	0.5	-0.565	0.002	0.356	0.041	16.275	0.981	0.042	0.950	0.044	0.804	0.051	0.849	0.049	0.495	0.074	0.029	0.024	0.021
	0.1	0.6	-0.574	0.002	0.349	0.050	16.196	0.981	0.051	0.954	0.052	0.805	0.062	0.848	0.059	0.490	0.095	0.029	0.024	0.022
	0.1	0.7	-0.620	-0.001	0.334	0.059	16.265	0.980	0.061	0.949	0.062	0.803	0.074	0.846	0.070	0.494	0.115	0.028	0.021	0.019
	0.1	0.8	-0.656	-0.004	0.310	0.056	16.312	0.979	0.057	0.950	0.058	0.802	0.069	0.844	0.066	0.497	0.111	0.027	0.019	0.019
	0.1	0.9	-0.673	-0.009	0.303	0.067	15.897	0.981	0.068	0.954	0.070	0.802	0.084	0.844	0.079	0.476	0.141	0.027	0.019	0.021

The size of the Wald test for all the samples, that is without conditioning on the F-statistic exceeding any critical value at all, doesn't seem too badly distorted – the worst actual size observed predictably corresponds to the case with the strongest simulated degree of endogeneity of x and the instrument that is uncorrelated with the highly endogenous x. In this case in the 26% of all the samples' Wald tests rejected the null hypothesis that β is equal to its true value of zero. However, it's worth noting that for the vast majority of cases even when the instrument is uncorrelated with the endogenous regressor, the actual size of the test doesn't exceed its nominal size and in most cases the test is actually undersized. Does this mean that no matter how weak the instrument is the inference based on the Wald test is reliable? This question is ambiguous; on one hand it looks like the correct null hypothesis will not be rejected most of the time, but at the same time the distribution of $\hat{\beta}$ is centered around the wrong value and the spread is so large, which suggests that the power of the test would be very low and the test will not be able to reject a large range of hypothesized wrong βs.

What happens when we start restricting the samples to the samples with the F-stat above the critical values? If we look at panel A columns (12) and (13), which report the fraction of the samples with the F-stat above 5.53 (the critical value for the 25% maximal size of the Wald test), we can see that the actual size for the samples with $\sigma_{xu} = 0.9$ is 99%, which means that 99% of the time the true β will be rejected. And as we keep raising the cut-off for the first stage F-stat, the size properties keep getting worse and converge to rejecting the true β 100% of the time for the F-stat above 9. However, the fraction of the samples with a high F-stat is very small, so it's quite unlikely that a researcher will find a first stage F-stat above the critical value when the instrument is irrelevant.

The situation becomes more problematic in panel B, where about 10% of the samples have F-statistics above 5.53 and it can be seen from column 13 that as the correlation between x and the error term increases the size of the test gets more and more distorted. For the samples with $\sigma_{x\varepsilon} = 0.9$ the rejection rate of the true β is about 77%, which contradicts Stock and Yogo's prediction of the 25% maximal size. The distortions get worse as the cut-off F-stat increases as well and the actual size converges to 100% for the 10% maximal bias critical value – an F-stat of 16.38.

As true correlation between the instrument and the regressor rises, in panels C and D the actual size of the conditional samples doesn't get as extreme as in panles A and B.

Nevertheless, the pattern persists – samples with highly endogenous x and high F-statistics consistently have poorer size properties. It can be seen in panel D, when the degree of endogeneity is low, the rejection rate for all the samples is very similar to that of the samples with higher F-statistic, but when the correlation between x and the error term goes up, conditional samples show large test size distortions. For example, when the correlation $\sigma_{x\epsilon} = 0.8$ the rejection of the true β for all the samples is about 8%. If we restrict the samples to only those that pass Stock and Yogo's 10% maximal size critical value (F-stat>16.38), the actual size of the test goes up to about 30%. For the $\sigma_{x\epsilon} = 0.9$, the rejection rate is above 35%, which seems huge compared to 8% for all the samples. It's worth noting that the fraction of the samples that pass the test is non-trivial –about one fifth of all the samples.

The correlation of 0.3 between the instrument and the regressor is fairly high, so why is the actual size of the test so different from the nominal size? One plausible explanation was presented by Hall et al. (1996); they suggested that when the correlation between the regressor and the error term is high and the instrument is strongly relevant, z inevitably explains some part of the variation in x that is endogenous, which basically means that the instrument is also endogenous. It seems reasonable to assume that as the correlation between x and z rises, the degree of endogeneity of z would also go up.

Columns 18-20 describe the behaviour of the median-squared error for all the samples and for the "conditional" samples, which at high levels of endogeneity of x seems consistent with what we saw earlier with the test size behaviour. The median squared error, which reflects the accuracy of prediction by the 2SLS estimator, is lower for the samples with high F-stats when the degree of endogeneity of x is low, but as the correlation between x and ε increases, the accuracy of prediction for the sample with higher F-stat deteriorates and we can see, for example, in panel D when $\sigma_{x\epsilon} = 0.8$, the median-squared error for the samples with F-statistics above 16.38 is double of that in the samples with the F-statistics above 8.96.

The results of these simulations may at first seem puzzling. It seems to be the case that the researcher is strictly worse off if he decides to winnow out instruments based on the pre-test suggested by Stock and Yogo as opposed to just using a random instrument without conducting a pre-test. The answer to this question can be yes or no, depending on how the researcher thinks about the pre-test. Consider the case when the researcher has an instrument that yields a first stage F-stat of about 5.5 (which corresponds to the critical value for the 25% maximal bias). What does this say about the correlation between the endogenous regressor and the instrument? From table 2.1 we can see that median F-stat for $\sigma_{xz} = 0.2$ is around 4 and the median F-stat for $\sigma_{xz} = 0.3$ is about 9; this suggests that the true correlation between x and z in the case where the F-statistic of 5.5 is obtained is likely to be somewhere between 0.2 and 0.3, which is fairly high and we can see in panel C, that the test size doesn't exceed "predicted" by Stock and Yogo size of 25% if we restrict the sample to those with the F-statistic above 5.53. This suggests that the pre-test "works" if the F-statistic the researcher obtained is not an unlikely realization for the true correlation of x and z.

How does the researcher know whether the F-stat obtained is somehow representative of the inherent relationship between the regressor and the instrument or just an "accident" from the extreme right tail of the F-distribution? In this case other information such as a previously conducted test or strong theoretical justification becomes relevant and can make a big difference. This pre-test should not be taken as a rule; it should certainly be used but with caution and with a fair share of skepticism. If the researcher suspects that the instrument at hand is a weak instrument, then the fact that the high F-statistic is obtained shouldn't trick him into thinking that his 2SLS results will be reliable, if anything, he should be more suspicious of his results if the F-stat was unexpectedly high. In this case, some form of the weak IV robust inference will be a better choice. However, if the researcher believes that his x is endogenous, but the degree of endogeneity is low, it seems that conditioning the choice of the instrument on the high value of the F-stat might be beneficial – the size distortions of the conditioned samples are similar to the non-conditioned samples, but there are definite gains in terms of accuracy of the estimates - the median-squared error is notably lower in the conditioned samples when the degree of endogeneity is not high.

Tables 2.2 and 2.3 support the conclusions drawn from table 2.1 – there seem to be no test size improvements from conditioning the choice of the instruments on higher values of the first stage F-statistic, but at lower levels of endogeneity of x, the median-squared error is significantly lower in the conditioned samples. For example, see table 2.3, panel B, $\sigma_{x\epsilon} = 0.1$ - the median-squared error for all the sample is 13 times larger than in the samples with F-statistics above 8.96 and 26 times larger than in the samples with the F-statistic above 16.38.

2.8 A possible solution to the pre-test bias.

Is there anything that can be done to avoid the pre-test bias? One possible solution to this problem, as noted by Hall et al. (1996) and Shea (1997) can be found in the Angrist and Krueger (1995) paper, where they critically assessed Angrist and Krueger (1991) results and suggested an estimator that unlike the 2SLS estimator would not be biased towards the OLS estimator. They suggested randomly splitting the sample into two subsamples, using one subsample to estimate the parameters of the reduced form equation and the second subsample to estimate the structural equation, but deploying the parameters from the first subsample. They called the estimator a split-sample IV estimator (SSIV) and showed that unlike the 2SLS estimator, which is biased towards OLS, it is biased towards zero, regardless of the strength of the correlation between the errors in the structural and in the reduced form equations. In the same paper, Angrist and Krueger also derive a USSIV – a split sample estimator that is asymptotically unbiased, where the bias is corrected by regressing the endogenous regressor on the predicted values from the first stage IV regression.

Angrist and Krueger (1995) replicated Angrist and Krueger (1991) using the SSIV and USSIV and obtained estimates similar to those obtained from using 2SLS and OLS estimators. They also showed that SSIV and USSIV when used to reproduce Bound et al (1995) experiment with randomly generated quarter of birth instruments unlike 2SLS do not spuriously produce estimates similar to the ones obtain using the real data.

The next section describes tests to discover whether problems associated with the pre-test bias can be eliminated by using the SSIV estimator – the idea is that if one half of the sample is used for the pre-test and the other for the estimation, the pre-test is conducted based on "extra" information, not the same information that will be used later for the estimation, so there is no sequential inference problem.

Design of the Monte Carlo simulations is very similar to the one described in the earlier experiment, the differences include the following:

Case 4: As in the previous experiment (*case 1*) 100 observations are generated, but now only the first 50 observations are used to estimate the first stage regression coefficients

Π, which are recorded along with the first stage F-statistics. The estimated Πs are then used to construct the fitted values for x for the next 50 observation and those fitted values are used as instruments in the second stage regression to estimate βs.

In order to keep the F-statistic comparable across the experiment with the 2SLS and SSIV when the number of observations in the first stage regression in decreased by a half, the correlation σ_{xz} is set to vary from 0 to 0.56 as opposed to from 0 to 0.4, that is the range of correlations is scaled up by a factor of $\sqrt{2}$. For the results table, the samples are conditioned on the F-stat that was obtained from the first stage regression using only half of the sample and then the size statistics are produced from the second stage regressions that only deploy the second half of the observations.

Case 5: This estimation is carried out in the same manner as the *case 4* estimation - half of the sample is used for estimating the first stage parameters and the second half for estimating the structural equation, but using Π obtained from the first half of the sample. However, the correlation σ_{xz} is kept in the same range as in the earlier experiment with the 2SLS estimator *(case 1)*. The idea here is to try to answer a more realistic question – for a given strength of the relationship between the instrument and the endogenous regressor, is it worth sacrificing a part of the sample in the estimation of the structural equation to avoid the pre-test bias that comes from sequential inference?

Dufour and Jasiak (2001) used a split-sample technique in the context of the AR-type inference and experimented with splitting the sample in different proportions. They concluded that the tests that used a smaller fraction for the first stage regression and most of the observations to estimate the structural equation were more powerful – their preferred split was a 1 to 9 ratio.

Case 6: This experiment is similar to *case 5* experiment, however, the split of the observations analogously to Dufour and Jasiak (2001) is changed to 10 and 90, 10 observations are used for the first stage regression and 90 for the estimation of the parameters of the structural equation. The correlation in parallel to *case 5* is unchanged (the same as in *case 1* with the 2SLS estimator).

2.9 Discussion of the results (SSIV).

Let's consider table 2.4. At first glance the results look very similar to the table 2.1 results: in panels A, B and C the median F-statistics (column 7) are almost the same as in the 2SLS experiment, but they get slightly larger in panels D and E. As the correlation between the instrument and the regressor rises, the median of the distribution of estimated βs (column 4) is approaching the true value for any degree of endogeneity. Also, the fractions of the samples that fall in the corresponding intervals that are conditioned on passing a certain critical value for the F-statistic (columns 8,10,12,14,16) are almost identical.

What about the size properties? The actual size of the Wald test (with the 5% nominal size) for all the samples (column 6) behaves very similarly to that in the 2SLS estimation – the maximum size observed is 25.5% and it corresponds to the samples with $\sigma_{xz} = 0$ and $\sigma_{x\varepsilon} = 0.9$ (the maximum size in table 2.1 corresponds to the same values for σ_{xz} and $\sigma_{x\varepsilon}$ and amounts to 26.2%). Also, similarly to table 2.1, as the correlation between *x* and *z* increases, the maximum size converges to about 6%.

However, when we start conditioning the samples the situation is drastically different. Column 9 reports the size of the Wald test for the samples that have significant at 5% F-statistics. The maximum size in those samples in panel A is 27%, while the maximum size in table 2.1 in the corresponding panel was 97%.

About 2% of the samples in both table 2.1 and table 2.4 have an F-stat above 5.53, a critical value for the 25% maximal size of the Wald test. In both table 2.1 and table 2.4 the maximal size (column 13) corresponds to $\sigma_{xz} = 0$ and $\sigma_{x\varepsilon} = 0.9$ and reaches 99% in Table 1 and only 23% in table 2.4.

In table 2.1 we consistently observed larger distortions to the Wald test sizes when the samples were conditioned on different values of the F-stats. And the distortions were getting worse with the increase in the critical values, which was true for any correlation between z and x and any degree of endogeneity. The SSIV produces very different results – size distortions are much smaller (the size never exceeds 27% as opposed to 100% for the 2SLS) and the size mainly decreases when the F-statistic increases.

Table 2.4. Case 4: SSIV, n=100 (50/50), 1 endogenous regressor, 1 instrument.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$					Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95	Size of stat(β˚) (all samples)	Median F-stat	Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat(β˚)	Fraction of all samples	Size of tstat(β˚)	Fraction of all samples	Size of tstat(β˚)	Fraction of all samples	Size of tstat(β˚)	All samples	Fstat>8.96	Fstat>16.38
Panel A	0.0	0	-6.561	0.010	6.100	0.000	0.465	0.048	0.000	0.002	0.000	0.021	0.000	0.004	0.000	0.001	0.000	0.998	1.474	1.557
	0.0	0.1	-6.215	0.069	6.183	0.000	0.445	0.050	0.002	0.004	0.000	0.024	0.004	0.006	0.000	0.000	0.000	0.966	0.683	1.837
	0.0	0.2	-6.039	0.203	6.259	0.001	0.454	0.049	0.000	0.003	0.000	0.021	0.000	0.004	0.000	0.000		1.001	0.714	
	0.0	0.3	-5.636	0.296	6.490	0.003	0.457	0.049	0.000	0.002	0.000	0.024	0.000	0.004	0.000	0.000	0.000	0.982	0.758	0.230
	0.0	0.4	-5.160	0.381	5.904	0.005	0.451	0.049	0.006	0.002	0.000	0.021	0.014	0.003	0.000	0.000	0.000	0.997	0.874	458.8
	0.0	0.5	-4.806	0.515	5.564	0.014	0.446	0.050	0.016	0.002	0.000	0.023	0.004	0.004	0.027	0.000	0.000	0.977	0.928	0.065
	0.0	0.6	-4.337	0.614	5.839	0.031	0.469	0.049	0.034	0.003	0.037	0.022	0.045	0.004	0.045	0.000	0.000	1.053	0.843	94.40
	0.0	0.7	-3.819	0.694	4.904	0.061	0.469	0.050	0.071	0.003	0.040	0.022	0.077	0.004	0.054	0.000		0.992	0.580	
	0.0	0.8	-3.180	0.790	4.628	0.124	0.461	0.048	0.160	0.003	0.120	0.021	0.136	0.004	0.146	0.000	0.000	0.999	1.019	0.759
	0.0	0.9	-1.860	0.901	3.655	0.255	0.462	0.050	0.272	0.002	0.217	0.021	0.228	0.005	0.224	0.000		1.008	1.077	

Table 2.4. Case 4: SSIV, n=100 (50/50), 1 endogenous regressor, 1 instrument (continued).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$			Size of stat(β˚) (all samples)	Median F-stat	Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95			Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat(β˚)	Fraction of all samples	Size of tstat(β˚)	Fraction of all samples	Size of tstat(β˚)	Fraction of all samples	Size of tstat(β˚)	All samples	Fstat> 8.96	Fstat> 16.38
Panel B	0.14	0	-3.682	-0.003	3.663	0.001	1.095	0.160	0.003	0.021	0.005	0.091	0.002	0.030	0.003	0.003	0.040	0.431	0.376	0.293
	0.14	0.1	-3.878	0.043	3.744	0.001	1.118	0.165	0.001	0.021	0.000	0.098	0.000	0.030	0.000	0.002	0.000	0.451	0.576	0.458
	0.14	0.2	-3.762	0.086	3.886	0.003	1.075	0.164	0.002	0.021	0.000	0.098	0.003	0.029	0.000	0.002	0.000	0.423	0.413	0.226
	0.14	0.3	-3.866	0.116	3.885	0.007	1.100	0.163	0.010	0.016	0.012	0.094	0.010	0.026	0.011	0.002	0.000	0.423	0.424	0.329
	0.14	0.4	-3.698	0.150	3.880	0.013	1.120	0.166	0.010	0.021	0.014	0.096	0.009	0.030	0.013	0.003	0.040	0.418	0.443	0.446
	0.14	0.5	-3.779	0.208	3.846	0.024	1.088	0.164	0.023	0.022	0.018	0.097	0.025	0.031	0.022	0.003	0.000	0.427	0.339	0.286
	0.14	0.6	-3.885	0.228	3.861	0.046	1.092	0.156	0.046	0.020	0.060	0.089	0.047	0.027	0.051	0.003	0.071	0.414	0.427	0.642
	0.14	0.7	-3.889	0.280	4.086	0.073	1.059	0.156	0.072	0.020	0.089	0.092	0.076	0.030	0.103	0.002	0.045	0.408	0.403	0.934
	0.14	0.8	-3.968	0.296	3.956	0.116	1.106	0.165	0.104	0.021	0.092	0.097	0.100	0.028	0.086	0.002	0.208	0.368	0.383	0.292
	0.14	0.9	-3.922	0.303	4.456	0.164	1.115	0.167	0.165	0.021	0.135	0.096	0.161	0.029	0.142	0.002	0.050	0.328	0.281	0.133
Panel C	0.28	0	-1.262	0.002	1.195	0.006	4.203	0.515	0.006	0.151	0.007	0.384	0.007	0.192	0.008	0.035	0.008	0.123	0.123	0.119
	0.28	0.1	-1.302	0.010	1.125	0.008	4.193	0.514	0.007	0.150	0.009	0.387	0.007	0.188	0.008	0.035	0.011	0.129	0.120	0.104
	0.28	0.2	-1.511	0.004	1.074	0.010	4.180	0.511	0.010	0.153	0.007	0.382	0.009	0.191	0.006	0.035	0.014	0.125	0.124	0.114
	0.28	0.3	-1.620	0.006	1.031	0.016	4.253	0.518	0.015	0.155	0.014	0.393	0.017	0.197	0.014	0.038	0.008	0.122	0.121	0.124
	0.28	0.4	-1.649	0.020	0.896	0.027	4.213	0.519	0.025	0.151	0.023	0.387	0.026	0.190	0.025	0.037	0.033	0.125	0.128	0.133
	0.28	0.5	-1.913	0.015	0.838	0.039	4.181	0.512	0.039	0.148	0.041	0.384	0.041	0.189	0.043	0.034	0.032	0.125	0.127	0.121
	0.28	0.6	-2.024	0.041	0.756	0.057	4.145	0.508	0.059	0.142	0.055	0.379	0.059	0.179	0.058	0.032	0.063	0.112	0.110	0.108
	0.28	0.7	-2.119	0.032	0.666	0.071	4.039	0.500	0.070	0.141	0.065	0.377	0.067	0.177	0.067	0.033	0.055	0.109	0.109	0.112
	0.28	0.8	-2.326	0.033	0.592	0.090	4.234	0.517	0.089	0.157	0.085	0.389	0.087	0.195	0.083	0.035	0.078	0.107	0.111	0.123
	0.28	0.9	-2.389	0.026	0.536	0.100	4.091	0.504	0.098	0.148	0.101	0.380	0.100	0.186	0.099	0.032	0.089	0.102	0.098	0.084

Table 2.4. Case 4: SSIV, n=100 (50/50), 1 endogenous regressor, 1 instrument (continued).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$					Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95	Size of stat(β*) (all samples)	Median F-stat	Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat(β*)	Fraction of all samples	Size of tstat(β*)	Fraction of all samples	Size of tstat(β*)	Fraction of all samples	Size of tstat(β*)	All samples	Fstat> 8.96	Fstat> 16.38
Panel D	0.42	0	-0.649	0.003	0.633	0.016	10.556	0.873	0.016	0.530	0.018	0.795	0.017	0.589	0.018	0.242	0.021	0.053	0.054	0.054
	0.42	0.1	-0.702	-0.007	0.596	0.018	10.606	0.871	0.018	0.533	0.017	0.794	0.018	0.593	0.017	0.238	0.015	0.053	0.053	0.053
	0.42	0.2	-0.734	-0.004	0.568	0.022	10.559	0.872	0.022	0.531	0.022	0.800	0.022	0.591	0.023	0.242	0.020	0.054	0.055	0.056
	0.42	0.3	-0.770	0.005	0.532	0.032	10.242	0.868	0.032	0.514	0.030	0.782	0.031	0.578	0.031	0.232	0.030	0.054	0.054	0.057
	0.42	0.4	-0.833	0.003	0.512	0.038	10.520	0.872	0.037	0.529	0.038	0.795	0.038	0.584	0.038	0.240	0.038	0.053	0.055	0.054
	0.42	0.5	-0.866	-0.009	0.475	0.046	10.572	0.872	0.046	0.534	0.046	0.800	0.047	0.592	0.046	0.249	0.045	0.051	0.052	0.055
	0.42	0.6	-0.949	-0.003	0.444	0.056	10.405	0.875	0.055	0.523	0.056	0.795	0.056	0.582	0.055	0.246	0.048	0.053	0.051	0.054
	0.42	0.7	-0.984	-0.004	0.424	0.067	10.488	0.874	0.067	0.528	0.063	0.796	0.068	0.585	0.066	0.239	0.059	0.051	0.050	0.050
	0.42	0.8	-1.053	-0.002	0.394	0.076	10.571	0.874	0.079	0.529	0.078	0.795	0.078	0.592	0.079	0.243	0.078	0.050	0.049	0.048
	0.42	0.9	-1.146	-0.004	0.371	0.080	10.712	0.875	0.080	0.539	0.081	0.794	0.080	0.599	0.080	0.239	0.075	0.047	0.048	0.045
Panel E	0.56	0	-0.448	0.006	0.455	0.032	22.744	0.991	0.032	0.911	0.032	0.981	0.032	0.931	0.032	0.724	0.032	0.029	0.029	0.029
	0.56	0.1	-0.464	0.001	0.419	0.030	22.191	0.993	0.030	0.913	0.031	0.981	0.030	0.934	0.031	0.721	0.030	0.029	0.029	0.029
	0.56	0.2	-0.495	-0.001	0.410	0.033	22.384	0.991	0.033	0.916	0.032	0.981	0.032	0.938	0.033	0.723	0.032	0.028	0.028	0.028
	0.56	0.3	-0.528	-0.002	0.389	0.038	22.227	0.992	0.039	0.913	0.039	0.982	0.039	0.935	0.039	0.714	0.041	0.029	0.030	0.030
	0.56	0.4	-0.536	-0.001	0.370	0.043	22.706	0.992	0.043	0.912	0.043	0.981	0.043	0.935	0.043	0.722	0.041	0.028	0.028	0.028
	0.56	0.5	-0.549	0.000	0.359	0.049	22.770	0.993	0.048	0.915	0.048	0.985	0.048	0.940	0.048	0.725	0.048	0.029	0.028	0.029
	0.56	0.6	-0.575	0.002	0.345	0.053	22.419	0.992	0.053	0.911	0.053	0.981	0.053	0.935	0.053	0.719	0.053	0.028	0.028	0.028
	0.56	0.7	-0.596	0.001	0.333	0.059	22.958	0.993	0.059	0.919	0.059	0.983	0.059	0.940	0.058	0.732	0.058	0.029	0.030	0.030
	0.56	0.8	-0.641	0.003	0.313	0.063	22.579	0.994	0.063	0.916	0.063	0.984	0.063	0.936	0.063	0.721	0.061	0.028	0.028	0.027
	0.56	0	-0.448	0.006	0.455	0.032	22.744	0.991	0.032	0.911	0.032	0.981	0.032	0.931	0.032	0.724	0.032	0.029	0.029	0.029

Table 2.5. Case 5: SSIV, n=100 (50/50), 1 endogenous regressor, 1 instrument.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$			Size of stat($\hat{\beta}$) (all samples)	Median F-stat Size of tstat(β)	Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
								Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Fstat> 8.96	Fstat> 16.38
	σ_{xz}	σ_{xu}	0.05	0.50	0.95															
Panel A	0.0	0	-6.561	0.010	6.100	0.000	0.465	0.048	0.000	0.002	0.000	0.021	0.000	0.004	0.000	0.001	0.000	0.998	1.474	1.557
	0.0	0.1	-6.215	0.069	6.183	0.000	0.445	0.050	0.002	0.004	0.000	0.024	0.004	0.006	0.000	0.000	0.000	0.966	0.683	1.837
	0.0	0.2	-6.039	0.203	6.259	0.001	0.454	0.049	0.000	0.003	0.000	0.021	0.000	0.004	0.000	0.000		1.001	0.714	0.000
	0.0	0.3	-5.636	0.296	6.490	0.003	0.457	0.049	0.000	0.002	0.000	0.024	0.000	0.004	0.000	0.000	0.000	0.982	0.758	0.230
	0.0	0.4	-5.160	0.381	5.904	0.005	0.451	0.049	0.006	0.002	0.000	0.021	0.014	0.003	0.000	0.000	0.000	0.997	0.874	458.8
	0.0	0.5	-4.806	0.515	5.564	0.014	0.446	0.050	0.016	0.002	0.000	0.023	0.004	0.004	0.027	0.000	0.000	0.977	0.928	0.065
	0.0	0.6	-4.337	0.614	5.839	0.031	0.469	0.049	0.034	0.003	0.037	0.022	0.045	0.004	0.045	0.000	0.000	1.053	0.843	94.39
	0.0	0.7	-3.819	0.694	4.904	0.061	0.469	0.050	0.071	0.003	0.040	0.022	0.077	0.004	0.054	0.000		0.992	0.580	0.000
	0.0	0.8	-3.180	0.790	4.628	0.124	0.461	0.048	0.160	0.003	0.120	0.021	0.136	0.004	0.146	0.000	0.000	0.999	1.019	0.759
	0.0	0.9	-1.860	0.901	3.655	0.255	0.462	0.050	0.272	0.002	0.217	0.021	0.228	0.005	0.224	0.000		1.008	1.077	0.000

Table 2.5. Case 5: SSIV, n=100 (50/50), 1 endogenous regressor, 1 instrument.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$			Size of stat(β̂) (all samples)	Median F-stat	Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95			Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat(β̂)	Fraction of all samples	Size of tstat(β̂)	Fraction of all samples	Size of tstat(β̂)	Fraction of all samples	Size of tstat(β̂)	All samples	Fstat> 8.96	Fstat> 16.38
Panel B	0.1	0	-4.549	0.002	4.771	0.001	0.730	0.103	0.002	0.010	0.000	0.055	0.002	0.015	0.000	0.001	0.000	0.622	0.671	1.601
	0.1	0.1	-5.210	0.072	4.892	0.001	0.751	0.107	0.000	0.009	0.000	0.058	0.000	0.015	0.000	0.001	0.000	0.639	0.558	0.382
	0.1	0.2	-4.552	0.143	5.106	0.002	0.724	0.107	0.002	0.010	0.000	0.056	0.002	0.015	0.000	0.001	0.000	0.625	0.889	0.110
	0.1	0.3	-4.731	0.174	4.914	0.005	0.760	0.105	0.009	0.008	0.024	0.054	0.013	0.013	0.016	0.001	0.000	0.615	1.361	0.092
	0.1	0.4	-4.563	0.227	4.773	0.010	0.765	0.107	0.006	0.011	0.009	0.061	0.008	0.016	0.006	0.001	0.000	0.614	0.645	0.255
	0.1	0.5	-4.661	0.311	4.879	0.020	0.731	0.107	0.021	0.011	0.000	0.059	0.020	0.015	0.020	0.001	0.000	0.626	0.611	1.602
	0.1	0.6	-4.610	0.341	4.743	0.039	0.723	0.101	0.032	0.010	0.038	0.052	0.035	0.015	0.046	0.001	0.000	0.621	0.699	1.247
	0.1	0.7	-4.113	0.430	4.826	0.070	0.713	0.102	0.071	0.009	0.077	0.058	0.081	0.015	0.075	0.001	0.000	0.603	0.701	1.829
	0.1	0.8	-3.602	0.471	4.576	0.121	0.750	0.106	0.102	0.011	0.109	0.059	0.096	0.016	0.119	0.001	0.167	0.584	0.596	0.303
	0.1	0.9	-3.376	0.495	4.602	0.202	0.770	0.107	0.206	0.010	0.158	0.057	0.215	0.015	0.192	0.001	0.167	0.513	0.488	0.271
Panel C	0.2	0	-2.335	0.009	2.403	0.003	2.084	0.290	0.002	0.053	0.004	0.193	0.003	0.073	0.003	0.008	0.000	0.241	0.225	0.180
	0.2	0.1	-2.532	0.027	2.283	0.003	2.079	0.290	0.003	0.053	0.004	0.190	0.004	0.073	0.004	0.009	0.000	0.252	0.220	0.146
	0.2	0.2	-2.512	0.040	2.323	0.005	2.070	0.285	0.005	0.049	0.006	0.190	0.004	0.066	0.005	0.009	0.022	0.246	0.239	0.159
	0.2	0.3	-2.566	0.047	2.181	0.010	2.110	0.297	0.008	0.055	0.005	0.198	0.006	0.074	0.008	0.009	0.000	0.235	0.234	0.213
	0.2	0.4	-2.841	0.076	2.117	0.020	2.109	0.292	0.019	0.054	0.020	0.191	0.021	0.074	0.022	0.007	0.028	0.241	0.239	0.419
	0.2	0.5	-2.712	0.085	2.197	0.030	2.072	0.292	0.031	0.051	0.029	0.190	0.034	0.069	0.030	0.008	0.013	0.241	0.225	0.202
	0.2	0.6	-2.850	0.121	2.417	0.054	2.018	0.280	0.057	0.048	0.048	0.179	0.051	0.066	0.045	0.008	0.062	0.215	0.186	0.129
	0.2	0.7	-3.263	0.115	2.185	0.074	1.963	0.279	0.071	0.049	0.062	0.177	0.066	0.066	0.065	0.007	0.042	0.210	0.218	0.224
	0.2	0.8	-3.042	0.130	2.767	0.106	2.123	0.294	0.105	0.052	0.092	0.197	0.106	0.073	0.099	0.007	0.081	0.198	0.180	0.163
	0.2	0.9	-3.443	0.121	3.345	0.127	2.009	0.286	0.125	0.051	0.131	0.190	0.132	0.068	0.137	0.008	0.143	0.186	0.176	0.137

Table 2.5. Case 5: SSIV, n=100 (50/50), 1 endogenous regressor, 1 instrument.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$			Size of stat($\hat{\beta}$) (all samples)	Median F-stat	Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
								Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	All samples	Fstat> 8.96	Fstat> 16.38
	σ_{xz}	σ_{xu}	0.05	0.50	0.95															
Panel D	0.3	0	-1.124	0.008	1.109	0.005	4.870	0.572	0.006	0.190	0.007	0.445	0.006	0.236	0.007	0.049	0.012	0.109	0.110	0.124
	0.3	0.1	-1.167	-0.005	1.011	0.008	4.928	0.578	0.008	0.185	0.005	0.450	0.009	0.233	0.007	0.048	0.008	0.109	0.113	0.110
	0.3	0.2	-1.275	0.001	0.949	0.012	4.902	0.578	0.013	0.196	0.009	0.445	0.012	0.235	0.010	0.048	0.006	0.111	0.105	0.088
	0.3	0.3	-1.430	0.015	0.859	0.022	4.707	0.561	0.019	0.182	0.020	0.435	0.019	0.226	0.018	0.046	0.017	0.109	0.114	0.135
	0.3	0.4	-1.504	0.021	0.828	0.029	4.837	0.571	0.028	0.188	0.030	0.441	0.028	0.231	0.029	0.048	0.034	0.107	0.110	0.115
	0.3	0.5	-1.590	-0.001	0.726	0.041	4.926	0.580	0.042	0.194	0.042	0.450	0.041	0.239	0.043	0.053	0.040	0.103	0.107	0.101
	0.3	0.6	-1.790	0.018	0.658	0.057	4.809	0.572	0.055	0.195	0.047	0.444	0.054	0.236	0.052	0.049	0.039	0.102	0.102	0.108
	0.3	0.7	-1.923	0.008	0.606	0.076	4.816	0.572	0.074	0.188	0.069	0.441	0.072	0.233	0.070	0.048	0.065	0.098	0.093	0.105
	0.3	0.8	-2.131	0.017	0.546	0.088	4.867	0.572	0.089	0.189	0.089	0.445	0.086	0.233	0.089	0.050	0.070	0.095	0.091	0.073
	0.3	0.9	-2.142	0.010	0.496	0.095	4.966	0.581	0.092	0.189	0.084	0.452	0.094	0.233	0.090	0.048	0.092	0.090	0.094	0.088
Panel E	0.4	0	-0.697	0.007	0.705	0.016	9.523	0.834	0.015	0.473	0.014	0.748	0.015	0.535	0.014	0.195	0.017	0.059	0.058	0.058
	0.4	0.1	-0.733	0.001	0.636	0.016	9.266	0.835	0.017	0.455	0.015	0.743	0.016	0.520	0.016	0.183	0.016	0.059	0.056	0.059
	0.4	0.2	-0.796	0.000	0.611	0.019	9.333	0.843	0.019	0.461	0.020	0.748	0.019	0.523	0.018	0.196	0.017	0.058	0.056	0.058
	0.4	0.3	-0.858	0.003	0.571	0.028	9.252	0.838	0.029	0.458	0.033	0.743	0.030	0.519	0.031	0.187	0.038	0.059	0.060	0.058
	0.4	0.4	-0.912	0.000	0.535	0.037	9.463	0.836	0.037	0.472	0.036	0.743	0.036	0.530	0.036	0.192	0.036	0.056	0.055	0.056
	0.4	0.5	-0.977	0.000	0.495	0.049	9.478	0.840	0.048	0.467	0.046	0.747	0.047	0.531	0.048	0.197	0.043	0.057	0.056	0.051
	0.4	0.6	-1.000	0.007	0.475	0.056	9.317	0.836	0.057	0.459	0.059	0.744	0.056	0.524	0.058	0.187	0.061	0.054	0.055	0.059
	0.4	0.7	-1.084	0.008	0.440	0.067	9.639	0.846	0.066	0.477	0.066	0.759	0.066	0.534	0.067	0.194	0.068	0.057	0.056	0.054
	0.4	0.8	-1.165	0.005	0.406	0.073	9.615	0.842	0.072	0.475	0.070	0.748	0.071	0.538	0.070	0.199	0.065	0.054	0.055	0.055

Table 2.6. Case 6: SSIV, n=100 (10/90), 1 endogenous regressor, 1 instrument.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$					Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
						Size of stat(β̂') (all samples)	Median F-stat Size of tstat(β̂)	Fraction of all samples	Size of tstat(β̂')	Fraction of all samples	Size of tstat(β̂')	Fraction of all samples	Size of tstat(β̂)	Fraction of all samples	Size of tstat(β̂')	Fraction of all samples	Size of tstat(β̂')	Fraction of all samples	Fstat> 8.96	Fstat> 16.38
	σ_{xz}	σ_{xu}	0.05	0.50	0.95															
Panel A	0	0	-6.828	0.019	6.802	0.000	0.488	0.049	0.000	0.011	0.000	0.043	0.000	0.014	0.000	0.002	0.000	1.009	1.464	0.407
	0	0.1	-5.823	0.095	6.427	0.000	0.500	0.051	0.000	0.011	0.000	0.044	0.000	0.014	0.000	0.003	0.000	1.008	0.812	1.125
	0	0.2	-6.280	0.193	6.339	0.001	0.502	0.049	0.002	0.010	0.010	0.042	0.002	0.014	0.007	0.002	0.000	1.038	1.050	0.996
	0	0.3	-6.227	0.307	6.169	0.002	0.470	0.048	0.002	0.011	0.000	0.042	0.002	0.015	0.000	0.003	0.000	1.022	1.468	2.368
	0	0.4	-5.233	0.378	5.598	0.007	0.483	0.046	0.009	0.011	0.009	0.040	0.008	0.015	0.007	0.003	0.031	1.025	1.007	1.773
	0	0.5	-4.824	0.504	6.152	0.016	0.480	0.046	0.017	0.011	0.000	0.039	0.010	0.015	0.007	0.003	0.000	0.993	0.700	0.649
	0	0.6	-4.216	0.611	5.561	0.033	0.478	0.050	0.036	0.011	0.037	0.042	0.038	0.014	0.036	0.003	0.040	0.987	1.290	2.083
	0	0.7	-3.865	0.701	4.986	0.066	0.501	0.047	0.063	0.011	0.053	0.041	0.061	0.014	0.064	0.002	0.042	0.996	0.732	0.538
	0	0.8	-2.802	0.815	4.781	0.130	0.492	0.052	0.149	0.012	0.154	0.045	0.147	0.016	0.155	0.003	0.100	1.009	0.829	1.028
	0	0.9	-1.622	0.911	3.642	0.257	0.488	0.052	0.263	0.011	0.196	0.044	0.262	0.016	0.214	0.003	0.214	1.001	1.170	0.840

Table 2.6. Case 6: SSIV, n=100 (10/90), 1 endogenous regressor, 1 instrument (continued).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$			Size of stat($\hat{\beta}$) (all samples)	Median F-stat	Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95			Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	Fraction of all samples	Size of tstat($\hat{\beta}$)	All samples	Fstat> 8.96	Fstat> 16.38
Panel B	0.1	0	-4.258	-0.001	3.905	0.001	0.580	0.060	0.000	0.016	0.000	0.053	0.000	0.020	0.000	0.003	0.000	0.463	0.410	0.124
	0.1	0.1	-4.026	0.042	3.840	0.001	0.549	0.062	0.000	0.015	0.000	0.055	0.000	0.020	0.000	0.004	0.000	0.474	0.565	0.468
	0.1	0.2	-4.089	0.095	4.369	0.003	0.553	0.063	0.003	0.016	0.006	0.054	0.002	0.020	0.005	0.005	0.000	0.509	0.409	0.368
	0.1	0.3	-3.996	0.135	3.953	0.006	0.525	0.061	0.003	0.017	0.000	0.053	0.004	0.022	0.004	0.004	0.000	0.472	0.560	0.604
	0.1	0.4	-3.918	0.175	4.192	0.014	0.515	0.055	0.022	0.014	0.036	0.047	0.023	0.017	0.029	0.004	0.071	0.456	0.376	0.374
	0.1	0.5	-3.959	0.236	4.628	0.024	0.545	0.058	0.026	0.015	0.026	0.051	0.026	0.019	0.026	0.004	0.056	0.467	0.483	0.714
	0.1	0.6	-3.786	0.263	4.152	0.041	0.535	0.059	0.043	0.015	0.060	0.050	0.048	0.019	0.056	0.004	0.073	0.435	0.354	0.216
	0.1	0.7	-3.934	0.295	4.151	0.075	0.536	0.059	0.061	0.014	0.086	0.052	0.065	0.019	0.070	0.003	0.000	0.427	0.466	0.445
	0.1	0.8	-4.085	0.325	4.574	0.114	0.585	0.059	0.103	0.015	0.112	0.052	0.105	0.020	0.117	0.005	0.174	0.404	0.408	0.433
	0.1	0.9	-3.724	0.328	4.945	0.167	0.559	0.058	0.166	0.015	0.170	0.050	0.163	0.018	0.168	0.004	0.154	0.349	0.373	0.466
Panel C	0.2	0	-1.350	0.003	1.494	0.003	0.755	0.093	0.001	0.023	0.000	0.080	0.001	0.031	0.000	0.007	0.000	0.142	0.129	0.138
	0.2	0.1	-1.462	0.005	1.307	0.005	0.753	0.094	0.006	0.029	0.007	0.083	0.006	0.035	0.006	0.009	0.000	0.133	0.129	0.119
	0.2	0.2	-1.581	0.012	1.219	0.009	0.720	0.087	0.007	0.025	0.008	0.077	0.008	0.032	0.009	0.007	0.000	0.138	0.159	0.173
	0.2	0.3	-1.770	0.030	1.139	0.015	0.707	0.088	0.012	0.025	0.012	0.080	0.011	0.033	0.012	0.006	0.031	0.136	0.154	0.152
	0.2	0.4	-1.776	0.026	1.016	0.025	0.695	0.090	0.020	0.027	0.018	0.078	0.019	0.034	0.018	0.009	0.022	0.133	0.154	0.146
	0.2	0.5	-2.010	0.033	0.946	0.040	0.753	0.088	0.042	0.026	0.027	0.078	0.042	0.033	0.040	0.007	0.014	0.136	0.128	0.115
	0.2	0.6	-2.152	0.031	0.822	0.056	0.774	0.091	0.050	0.024	0.050	0.081	0.050	0.031	0.045	0.006	0.079	0.127	0.116	0.132
	0.2	0.7	-2.238	0.032	0.766	0.071	0.696	0.088	0.068	0.023	0.064	0.079	0.070	0.029	0.078	0.007	0.074	0.128	0.141	0.121
	0.2	0.8	-2.551	0.033	0.647	0.085	0.714	0.088	0.095	0.025	0.112	0.078	0.098	0.031	0.124	0.008	0.141	0.116	0.141	0.147
	0.2	0.9	-2.929	0.046	0.589	0.111	0.708	0.090	0.094	0.025	0.115	0.078	0.100	0.032	0.104	0.007	0.100	0.115	0.106	0.122

Table 2.6. Case 6: SSIV, n=100 (10/90), 1 endogenous regressor, 1 instrument (continued).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Estimated fractiles of $\hat{\beta}$			Size of stat(β˙) (all samples)	Median F-stat	Fstat significant at 5% level		Fstat>10		Fstat>5.53		Fstat>8.96		Fstat>16.38		Median-squared error, for the samples with		
	σ_{xz}	σ_{xu}	0.05	0.50	0.95			Fraction of all samples	Size of tstat(β)	Fraction of all samples	Size of tstat(β˙)	Fraction of all samples	Size of tstat(β˙)	Fraction of all samples	Size of tstat(β˙)	Fraction of all samples	Size of tstat(β˙)	All samples	Fstat> 8.96	Fstat> 16.38
Panel D	0.3	0	-0.710	0.003	0.697	0.011	1.120	0.139	0.010	0.042	0.010	0.123	0.010	0.053	0.011	0.014	0.014	0.058	0.056	0.057
	0.3	0.1	-0.750	0.008	0.662	0.017	1.147	0.145	0.015	0.044	0.011	0.131	0.016	0.058	0.021	0.013	0.008	0.061	0.059	0.042
	0.3	0.2	-0.785	0.006	0.622	0.019	1.134	0.139	0.020	0.043	0.023	0.124	0.020	0.055	0.025	0.013	0.015	0.057	0.074	0.062
	0.3	0.3	-0.854	0.006	0.579	0.026	1.203	0.144	0.023	0.047	0.021	0.132	0.023	0.060	0.023	0.015	0.033	0.059	0.061	0.063
	0.3	0.4	-0.916	-0.005	0.534	0.033	1.133	0.147	0.037	0.046	0.048	0.129	0.036	0.056	0.043	0.015	0.052	0.059	0.060	0.063
	0.3	0.5	-0.966	-0.001	0.505	0.045	1.133	0.147	0.046	0.046	0.046	0.133	0.048	0.059	0.044	0.015	0.048	0.057	0.066	0.076
	0.3	0.6	-1.036	0.008	0.464	0.053	1.182	0.149	0.055	0.046	0.058	0.134	0.055	0.057	0.053	0.013	0.052	0.055	0.042	0.038
	0.3	0.7	-1.149	-0.001	0.441	0.065	1.148	0.144	0.060	0.048	0.061	0.130	0.059	0.058	0.057	0.014	0.050	0.057	0.050	0.051
	0.3	0.8	-1.177	0.006	0.419	0.080	1.196	0.144	0.074	0.046	0.076	0.129	0.074	0.058	0.078	0.014	0.057	0.055	0.059	0.049
	0.3	0.9	-1.290	0.000	0.391	0.083	1.153	0.148	0.072	0.047	0.088	0.133	0.073	0.059	0.078	0.013	0.107	0.053	0.052	0.055
Panel E	0.4	0	-0.472	0.000	0.463	0.021	1.902	0.231	0.026	0.085	0.025	0.211	0.025	0.102	0.023	0.030	0.023	0.033	0.033	0.032
	0.4	0.1	-0.503	-0.002	0.455	0.026	1.985	0.238	0.024	0.085	0.021	0.218	0.023	0.102	0.021	0.031	0.016	0.033	0.034	0.041
	0.4	0.2	-0.540	0.004	0.439	0.028	1.970	0.236	0.029	0.090	0.029	0.219	0.028	0.110	0.029	0.030	0.033	0.033	0.034	0.034
	0.4	0.3	-0.553	0.003	0.406	0.030	1.984	0.231	0.027	0.085	0.025	0.214	0.029	0.105	0.027	0.031	0.020	0.033	0.030	0.029
	0.4	0.4	-0.575	0.000	0.402	0.039	1.954	0.235	0.039	0.086	0.046	0.213	0.039	0.104	0.042	0.029	0.045	0.032	0.032	0.031
	0.4	0.5	-0.597	0.001	0.378	0.046	1.934	0.231	0.050	0.084	0.054	0.211	0.052	0.104	0.049	0.032	0.034	0.033	0.035	0.033
	0.4	0.6	-0.628	0.005	0.363	0.055	1.978	0.239	0.054	0.093	0.069	0.221	0.055	0.112	0.064	0.031	0.074	0.033	0.031	0.028
	0.4	0.7	-0.680	0.003	0.341	0.059	1.872	0.231	0.061	0.086	0.056	0.211	0.060	0.105	0.054	0.030	0.072	0.033	0.033	0.036
	0.4	0.8	-0.707	-0.003	0.334	0.071	1.927	0.237	0.068	0.088	0.071	0.218	0.068	0.107	0.068	0.031	0.059	0.032	0.031	0.030
	0.4	0.9	-0.709	-0.002	0.317	0.072	2.046	0.246	0.073	0.091	0.076	0.224	0.074	0.112	0.070	0.034	0.080	0.033	0.032	0.030

As noted earlier, Angrist and Krueger (1995) showed that the SSIV, unlike the 2SLS, is always biased towards zero. By construction, the true β is my experiment is also equal to zero, so it was necessary to check if the test size improvements are specific to the case where true β is zero. As a robustness check a further experiment was conducted where the true β was set to one. The size of the Wald test when using SSIV behaved in the same way as in the experiment with the true β of zero, that is, large size distortions are not observed in the conditioned samples for any degree of correlation between the regressor and the error term regardless of the strength of the instruments.

It can be seen in table 2.4 that with just a couple of exceptions Stock and Yogo's (2005) critical values accurately predict the extent of the Wald test size distortions in the experiment with the SSIV estimator, which is good news. However, the SSIV brings another problem – the behavior of the median-squared error has changed and even in the samples where the degree of endogeneity is low the conditional samples do not yield lower median-square errors. In fact, the outcome is very sporadic – there is no clear pattern associated with the median-square error and conditional sampling; sometimes it's higher in the samples with higher F-stat sometimes it's lower. But mainly it is very similar for all the samples and for the conditioned samples and corresponds to the median-square error for all the samples in table 2.1, which means that there are no obvious gains in accuracy of the estimation of β in the samples where the first stage F-statistic is higher.

In table 2.5 we can see that the test size distortions are mainly inside the upper bounds determined by Stock and Yogo's (2005) critical values. This is similar to the results obtained in table 2.4, despite the fact that the fraction of the samples that pass different critical values in table 2.5 is significantly smaller. Also, unlike the results in table 1 and analogously to table 2.4, the median-squared error doesn't fall when the samples are restricted to those with higher corresponding F-statistics. In fact, the median-squared error in table 2.5 is about 1.5 - 2 times larger than the median-squared error in all samples in both table 2.1 and table 2.4, which suggests that opting for the SSIV estimation with the 50-50 split (as opposed to the 2SLS) would lead to a loss in accuracy of the prediction of β.

In table 2.6, where the 10-90 split is used, we can still see the benefits of the SSIV estimator - the size of the Wald test in the samples with higher F-statistics barely exceeds the maximal size derived by Stock and Yogo (2005) despite the fact that the number of observations in the first stage regression is very small and the fraction of the samples that pass different critical values is tiny compared to all previous cases. Since there are more observations used in the second stage regression, the distribution of the 2SLS estimates is narrower and the median-squared error is smaller compared to the corresponding median-squared errors in table 2.5. However, it is also clear that if the sample size is as small as described in this experiment, even when the correlation between the endogenous regressor and the instrument is fairly high (for example, see panel E), the F-statistics obtained will be very low, which would make the pre-test useless, because recognizing a strong instrument would be almost impossible.

2.10. Practitioner's guide to using Stock and Yogo's critical values as a pre-test.

In this chapter I have examined how Stock and Yogo's (2005) critical values perform as a pre-test while evaluating its effectiveness by means of the median-squared error. My findings have significant value for applied economisits that work with models that require IV estimation. In this section I provide practical advice regarding the use of Stock and Yogo's critical values as a pre-test, and explain how to avoid potential complications. The main relevant results of my research can be outlined in the following points:

1. While evaluating the strength of the instrument by comparing the first stage F-statistic with Stock and Yogo's (2005) critical values, researchers should be cautious. ***If the instrument is suspected to be weak and the degree of endogeneity is suspected to be high***, a high first stage F-statistic obtained is very likely to correspond to a large bias in the IV estimation, and significant size distortions to the Wald test on the coefficient on the endogenous regressor.
2. Thus, if the researcher has a strong prior belief that the ***instrument is only weakly correlated with the endogenous regressor and the degree of endogeneity is high,*** weak IV robust statistics, such as AR (1949), Kleibergen's (2002) Lagrange Multiplier statistic, etc., with distributions that do not depend on the value of the concentration parameter, provide a better alternative.
3. ***If the sample size is large***, another alternative is to use ***SSIV estimation***

procedure, which eliminates the pre-test bias problem, so Stock and Yogo's (2005) critical values "predict" Wald test size distortions fairly accurately. However, if ***the correlation between the instrument and the regressor is very low***, the precision of estimation will suffer, so the trade-off should be considered carefully.

4. ***If the degree of endogeneity of the regressor of interest is expected to be low,*** choosing instruments that "pass" certain critical values is beneficial – the accuracy of estimation is notably higher for the samples conditioned on the high F-statistic. The test size distortions are also low.

To summarise, when choosing instruments and prior to resorting to the pre-test, it is important to consider the degree of endogeneity and instrument relevance in your model based on previous knowledge and theoretical reasoning. Sequential testing is associated with the pre-test bias and I have shown that under certain circumstances the cost of that bias might be large. I recommend that the information provided in this section is taken into account when Stock and Yogo's (2005) critical values are used as a pre-test to determine whether the instruments are strong or weak.

2.11 Conclusion

Nagar (1959) was the first to point out that the bias of the 2SLS estimator increases as the correlation between the instrument and the endogenous regressor decreases, but for a long time this was only considered to be a problem in small samples. Later it was shown that even large samples could produce biased estimates if the instruments are sufficiently weak. There has been a lot of research since the 1990s on how to detect and overcome problems that arise in the presence of weak IV. Stock and Yogo (2005) suggested using non-conventional critical values for the first stage F-statistic and linked them to the maximal relative bias of the IV estimator and the maximal size of the Wald test.

In this chapter I have analyzed the performance of Stock and Yogo's (2005) critical values as a pre-test in order to winnow out weak instruments based on Hall et al.'s (1996) experiment. The innovative contribution of my paper is in applying median-squared error as a loss function to evaluate the effectiveness of the pre-test, and in exploring the performance of the pre-test in the context of the SSIV estimator.

The conclusions that can be drawn from the simulations suggest that:

1. While using the 2SLS estimator, there are adverse consequences which come from choosing instruments based on high F-statistics: the higher the F-statistic, the more the Wald test size is likely to be distorted. In particular, if the first stage F-statistic obtained is high but the researcher suspects that the underlying relationship between the endogenous regressor and the instrument is weak, the second-stage inference is likely to be misleading and the maximal Wald test size distortions will not fall within the bounds predicted by Stock and Yogo's (2005) critical values. In fact, the size distortions are likely to be huge – the actual size observed in the simulations reached up to 100%.

2. When the degree of endogeneity is low, however, the high first stage F-statistics correspond to a lower median-squared error. And for these low-endogeneity cases, the distortions to the Wald test size are comparatively small. Thus conditioning the choice of an instrument on the high value of the F-statistic could actually improve the accuracy of the estimation of β.

3. When the SSIV estimator is used, the pre-test bias problem is virtually eliminated and the test size distortions are mainly within the bounds predicted by Stock and Yogo (2005). However, as fewer observations are used for estimating the structural equation, the accuracy of the estimation of β decreases, so the pre-test in conjunction with the SSIV estimator should be used with caution when the sample size is small (or when the correlation between the instrument and the endogenous regressor is expected to be very low).

The implications of the pre-test bias when testing for the strength of the instruments can be serious and can compromise the reliability of an IV estimation; nevertheless, the test suggested by Stock and Yogo (2005) is still very useful if used correctly.

The practical advice that follows is that the researcher should not rely entirely on the statistics to select the set of instruments, but also on exogenous information available – for example: earlier theoretical or empirical research which indicates that the degree of endogeneity is either very low or very high. Kennedy (2003) notes that pre-testing, which is also called sequential estimation or data mining, is not well justified in the context of causal inference. In the context of instrument selection, Shea (1997) suggests that prior reasoning is very important and all screening tests should just be viewed as an additional check.

Another important point was made by Dufour (1997), who states that in the presence of weak instruments, the Wald test has the disadvantage of producing bounded confidence sets with probability one instead of producing infinite or close to infinite confidence sets when β is unidentifiable or almost unidentifiable. This suggests that the researcher should opt for an Anderson and Rubin's (1959) approach or its analogues if the instrument is thought to be weak. More generally, the traditional approach I have reviewed in this chapter uses an estimator that requires the rank condition to be satisfied. However, if the researcher does not have strong priors that the rank condition is satisfied in theory, an alternative approach that drops the rank condition would likely be more appropriate.

Chapter 3. Oil Rents and the Real Exchange Rate.[5]

[5] Joint work with Erkal Ersoy

3.1 Introduction.

Is there a resource curse, and if so how does it work? Could it be that an increase in resource wealth leads to a real appreciation that crowds out other exports? Does this happen in practice? The topic has been widely studied in an effort to explain the poor growth performance of some resource-rich countries since it seems counter-intuitive for natural resource abundance to lead to sluggish growth patterns. In addition to making the casual observation that most resource-rich countries have relatively low levels of GDP, Sachs and Warner (2001) have empirically illustrated that high resource intensity is correlated with slow growth. Other post-war growth studies have corroborated this finding (c.f. Sachs and Warner 1995, 1997). As discussed in Sala-i-Martin and Subramanian (2003), the case of Nigeria is particularly interesting because real per capita oil revenues rose from $33 in 1965 to $325 in 2000 with very little rise in real GDP per capita. Counter-examples, such as Norway and Australia, have led to a debate about the existence and significance of resource curse leading to the conclusion that other factors, such as corruption, account for most of the problems associated with slow economic development. In order to shed light on this debate, Sachs and Warner (1997) have implemented empirical models that include up to nine control variables—including corruption—and found that natural resource abundance still plays a key role in determining growth rates. In light of this finding, it has been postulated that exchange rates affect growth through hindrance of export-led growth. However, the existence of the resource curse is still controversial: some authors (Sachs and Warner 1999, Atkinson and Hamilton 2003) argue that it exists and is important, while others disagree (Mehlum et al. 2006, Brunnschweiler and Bulte, 2008). For a review of evidence in favour of the resource curse in oil, see Ross (2012).

The reason that the resource curse literature is relevant in a study of the relationship between resource rents and the real exchange rate is that one of the mechanisms by which resources are meant to curse a country is via appreciation of the real exchange rate. This phenomenon is known as the Dutch disease, and the key mechanism generally thought to underlie it is the Balassa-Samuelson (also called the Harrod-Balassa-Samuelson) hypothesis, which suggests that real appreciation is the result of relative productivity changes of the tradable vs. the non-tradable sector in resource rich economies. The idea is that since the price of tradables is set on international markets (but non-tradables are not), and since labour within a country can move freely between both sectors, a rise in the productivity of the tradables sector will suck labour out of the non-tradables sector; this will raise employment in the tradable sector (but not the price, which is set internationally) and it will also raise the price level within the non-tradable sector (wages must rise so that some workers are induced to stay behind rather than everybody leaving for the traeable sector at once).

At the intersection of the Balassa-Samuelson hypothesis, the Dutch disease, and the resource curse is the notion that a resource bonanza could cause a real appreciation which in turn would make other exports (e.g. manufactured exports) uncompetitive (this happened to the Netherlands after the discovery of large gas fields in the country in 1959, hence the name Dutch disease). If the exports which were crowded out by the real appreciation were in industries with large positive externalities (e.g. agglomeration economies, economies of scale, knowledge spillovers, etc.), then by crowding them out, a country may reduce its long-run growth rate, even though it is following its comparative advantage in the short run. And while the presence or absence of a real appreciation in response to an increase in resource rents would not conclusively prove or disprove the existence of the resource curse, the one or the other finding could strengthen the case (if the channel exists) or weaken it (if the channel does not exist, which suggests the resource curse must operate through other means such as corruption of the political system). So it is worth checking whether the real appreciation channel appears to be important in practice.

The chapter contributes to the general literature on the Balassa-Samuelson effect, but specifically focuses on the oil-producing countries. There are three distinctive features of the chapter - a new dataset on revenues and costs of the oil exporters that has been constructed from Wood Mackenzie's (WM) Global Economic Model (GEM) and has not been used in this context before; a diverse set of oil-exporting countries analysed; and the use of the pool mean group estimator by Pesaran et al. (1999), that restricts the long-run coefficients for all panels to be the same, but allows the short-run coefficients to vary by panel to try to identify whether the Balassa-Samuelson effect exists in the oil-exporting countries. These new data and techniques available allow us to quantify the long-run relationship between movements in real exchange rates that could be attributed to the changes in the oil rents per capita. Two findings are worth noting - the evidence in favour of the Balassa-Samuelson hypothesis was found for most countries in the sample, but the magnitudes of the effect are small. No evidence was found for the mechanism in the OPEC countries.

The chapter is organised as follows: section 3.2 reviews the Balassa-Samuelson (B-S) hypothesis and its transition mechanism, section 3.3 provides a literature review, section 3.4 describes the dataset in detail, section 3.5 focuses on the determining stationarity properties of the variables, section 3.6 describes the econometrics methods used and provides the results of the estimations, section 3.7 concludes the analysis.

3.2 The Balassa-Samuelson effect and the transmission mechanisms

One popular explanation for the resource curse has a crowding-out logic. If activity X drives growth and the extraction of natural resources crowds-out this activity, natural resources harm growth through the elimination of activity X. This activity could be in a manufacturing industry with positive externalities that would lead to improved efficiency and international competitiveness. Since natural resource exports dominate, however, other industries cannot compete in the global market and productivity-boosting spillovers are minimal. Therefore, if production and exports of natural resources lead to the appreciation of the domestic currency, domestic economic growth would be hurt. In addition, positive wealth shocks from the natural resource sector result in higher demand for non-traded goods and create excess demand for non-traded products driving up their prices. This rise in prices include input costs and wages which squeezes profits in traded activities, including manufacturing, that use the non-traded products as inputs but sell on the international market at relatively fixed prices. The decline in manufacturing then has ramification that slow down the growth process.

Harrod (1933), Balassa (1964) and Samuelson (1964) all independently pointed at precisely this phenomenon. They noted that countries with more productive labour in the tradable sector should have relatively higher prices in their non-tradable sector. This would then lead to a higher overall price level in countries with productive tradable sectors and indirectly to the appreciation of the currency. For instance, consider oil-exporting countries A and B. The former is similar to the United Arab Emirates and has a highly productive oil sector in which capital and labour input costs are low. The latter is similar to Kazakhstan where productivity in the oil sector is considerably lower and input costs are much higher. Exports of oil from both countries represent a high fraction of total exports and a large portion of their GDP. Assuming capital is perfectly mobile across sectors within and between countries, but labour is mobile only within the country and not internationally, we would expect a higher overall price level in country A than in B. The external mechanism through which this occurs can be explained as follows:

$P_T = XRAT \times P_T^*$	Law of one price holds for tradable goods only
$RER = \frac{XRAT}{P}$	Rodrik (2008) and MacDonald and Vieira (2010)
$W_T = P_T \times MPL_T \rightarrow P_T = \frac{W_T}{MPL_T}$ $W_N = P_N \times MPL_N \rightarrow P_N = \frac{W_N}{MPL_N}$	Workers are paid their marginal product
$W_N = W_T$	Workers can move freely between sectors
$P = P_T^{\alpha} \times P_N^{1-\alpha}$	Overall price level composition
$\frac{P_N}{P_T} = \frac{1}{P_T} \times \frac{W_N}{MPL_N} = \frac{1}{P_T} \times \frac{W_T}{MPL_N} = \frac{P_T}{P_T} \times \frac{MPL_T}{MPL_N} = \frac{MPL_T}{MPL_N}$ $MPL_T \uparrow \Longrightarrow P_N \uparrow \Longrightarrow P \uparrow \Longrightarrow RER \downarrow$	 Appreciation of the currency

Table 3.1. The B-S transition mechanism

In table 3.1 P_T is price of tradables, P_N price of non-tradables, P_T^* price of tradables abroad, XRAT nominal exchange rate, which is defined as the number of units of the domestic currency that buy one US dollar, P overall domestic price level, α is the share of tradables in the overall domestic price level, W_T wages in tradable sector, W_N wages in non-tradable sector, MPL_T marginal product of labour in tradable sector, MPL_N marginal product of labour in non-tradable sector and RER real exchange rate. Using our definition of the exchange rate, increases in the exchange rates mean depreciation of the domestic currency.

Since the marginal product of labour in country A is higher than that in country B, $\mathrm{MPL}_{\mathrm{T}}^{\mathrm{A}} > \mathrm{MPL}_{\mathrm{T}}^{\mathrm{B}}$, the price level of non-tradables will be higher leading to a higher overall price level in the country. This, in turn, drives the appreciation of the real exchange rate.

In fact, Balassa has made the observation that "the greater are the productivity differentials in the production of tradable goods between countries, the larger will be the differences in wages and in the prices of services and correspondingly the greater will be the gap between purchasing power parity and the equilibrium exchange rate" (Balassa, 1964).

3.3 Literature review.

After Balassa (1964) popularised the aforementioned notion, it was adopted not only in the exchange rate and resource curse literature, but also led to a new niche of its own. Sachs and Warner (1995, 1997, 1999, 2001) as well as Atkinson and Hamilton (2003), Sala-i-Martin and Subramanian (2003), Mehlum et al. (2006) and Brunnschweiler and Bulte (2008) are a few of the significant contributions to the resource curse literature. Within the B-S literature, time series and panel analyses largely support the B-S hypothesis, whereas initial cross-sectional analyses led to mixed results.

Balassa (1964) was the first one to attempt to verify the B-S hypothesis empirically by regressing PPP as a percentage of exchange rate on per capita GNP– he analysed 12 OECD countries in 1960 and found a significant relationship, which was interpreted as a confirmation for his proposition. His study gave rise to large cross-sectional literature, which includes De Vries (1968), Officer (1976), Clague (1986, 1988), etc. Most studies use the real exchange rate or PPP as the dependent variable and the explanatory variables include different productivity measures, such as GDP per capita, ratio of productivities, real income etc. and the control variables, such as openness to trade, trade balance, money growth and so on. The literature didn't provide conclusive results and different specifications yielded different outcomes.

The “second wave” of literature on the productivity bias hypothesis was focused on the country level analysis, i.e. time series analysis, which was meant to take into account country-specific circumstances that couldn’t have been accounted for in the cross-sectional studies. These studies include Hseih (1982), Rogoff (1992), Bahmani – Oskooee and Rhee (1996) and others. Different time series approaches were utilised, including Johansen approach, Engle-Granger, Dickey-Fuller tests, ARDL approach etc. These studies mainly supported the hypothesis. A comprehensive literature review can be found in Bahmani–Oskooee and Nasir (2005).

The most recent group of studies that analyse the existence of the B-S mechanism are predominantly based on panel econometric methods. Nonstationary panel methods are a fairly new field in econometrics, and these methods were only adopted in the B-S literature late 1990s and early 2000s. Prior to the 1990s, most studies tested individual countries for cointegration and proceed with conventional panel methods such as Seemingly Unrelated Regressions (SUR) or Fixed Effects (FE) estimations. De Gregorio et al. (1994) and Asea and Mendoza (1994) were amongst the first influential ones – both papers included the demand side variables in accordance with Rogoff (1992), and Asea and Mendoza (1994) also showed that the ratio of sectoral productivity per capita should be used in the context of the B-S hypothesis instead of level productivity per capita, however, most papers consequently continue to use level data as a proxy.

Many papers questioned the assumptions of the model and the validity of the whole model versus parts of the model. For instance, Egert et al. (2003) implemented panel cointegration analysis to study nine Central and Eastern European countries using quarterly average labour productivity data over the period 1995 to 2000. Although their conclusion suggested strong evidence in favour of the B-S effect, the authors noted that only part of the phenomenon is being captured. They argued that the increase in price level could also be explained by increasing quality of the goods, which was not captured by the CPI index. Faria and Leon-Ledesma (2003) tested for evidence of the B-S effect in the long run using relative real output per capita levels as a proxy for relative labour productivity among four countries—Germany, Japan, the UK, and the USA—for the period 1960 to 1996. They implemented models using levels and first differences, but neither pointed to a significant long-run relationship between the price level ratios and output ratios. However they suggested that the rejection of B-S effect doesn't necessarily mean that PPP hypothesis holds - the investigation of the differenced output ratios suggested that causality exists, but that causality goes from price ratios to output ratios, which violates the assumptions of PPP. On the contrary, Choudhri and Khan's (2005) analysis of 16 developing countries with different income levels over the period 1976 to 1994 illustrated the existence of a long-run relationship between the countries' productivity differentials and their real exchange rates. However, the strength of the relationship is sensitive to variation in income levels and the authors argue that terms of trade also have an influence on real exchange rate.

Garcia-Solanes at el. (2008) extended the work done by Egert et al. (2003) and similarly to Asea and Mendoza (1994) found that the internal transmission mechanism (an increase in the overall price level in response to the increase in productivity in the tradable sector) holds, but that the appreciation of the real exchange rate cannot be fully attributed to productivity differentials. Their work involved six new EU countries of that year (Czech Republic, Estonia, Latvia, Lithuania, Poland and Slovak Republic) and six other countries from EU-15 (Finland, France, Italy, Netherlands, Spain and Sweden) using data from 1995 through 2004. They suggested that the reason that external transition mechanism is not fulfilled is the fact that PPP in tradable sector doesn't hold. However, Garcia-Solanes and Torrejon-Flores (2009), showed that both internal and external mechanisms work in their analysis of 16 Latin American countries conversely to 16 OECD countries, where only internal mechanism was confirmed. Drine and Rault (2002) in their analysis of 6 Asian countries using panel cointegration techniques also questioned the assumptions of the B-S hypothesis and provided evidence that two assumptions of the model – PPP for tradable goods and the relationship between prices of nontradables and the real exchange rate can be violated, which would explain rejection of the B-S hypothesis in empirical work.

Chong et al. (2012) evaluate the adjustment of the real exchange rate to its long-run equilibrium for 21 OECD countries and confirm that the B-S effect is not just an essential component of the equilibrium, but the size of the B-S effect varies by country and influences the speed of adjustment of the real exchange rate to the equilibrium after a shock.

Chinn (2000) estimated a panel error correction model and found some evidence in support of the productivity bias hypothesis in five East Asian countries. He also investigated effects of government spending and real oil prices on the real exchange rates and found that contrary to Chinn (1997) study of 14 OECD countries government spending didn't exhibit significant effect and the oil price was significant for only three countries in his sample, one oil exporter and two oil importers, with the predictable sign – appreciation of the currency for oil exporter (Indonesia) and depreciation for the other two (Japan and Korea).

Despite extensive coverage of the B-S hypothesis in the literature, there has not been much research that would focus on the oil-producing countries or even developing countries – most research concerns OECD countries regardless of the finding that B-S effect is more likely to be present in poorer countries. In the last decade transitional economies have gained attention, including oil exporters such as Russia and Kazakhstan, however manifestation of the productivity bias hypothesis through the oil-producing sector hasn't been explored despite its direct relevance – oil producers are mainly countries with large shares of primary exports.

One of the papers that investigates this relationship is Korhonen and Juurikkala (2009). They are mainly interested in the effects of the oil price on the real exchange rates of the OPEC countries, however they also use GDP per capita as a measure of productivity. They reject the B-S hypothesis for their sample, while finding strong and significant relationship between the oil price and the real exchange rates of those countries. They also emphasize that it would be useful to analyse all oil-exporting countries – members and non-members of OPEC - in one panel framework as similar results were found for non-OPEC countries, such as Russia (Kalcheva and Oomes (2007)), and OPEC Venzuela (Zalduendo (2006)).

Another paper that included Russia among other countries (South Eastern Europe, Ukraine and Turkey) is Egert (2005). He analysed the exchange rate behaviour in the transition economies with undervalued currencies (nominal exchange rate below the PPP exchange rate) and concluded that the Balassa-Samuelson effect isn't strong. In particular, Egert notes that the assumptions of B-S, such as that an increase in productivity causes relative price increases, don't seem to hold. A particularly interesting result of the paper is that oil revenues do not prove to be important for the exchange rate fluctuations in Russia. On the contrary, Egert et al. (2006) provide an extensive overview of the exchange rate behaviour in the 14 transition economies and despite stating that the movements in the exchange rate should not be attributed to the B-S mechanism, they note that Russia and Kazakhstan (both present in our dataset) are negatively affected by the Dutch Disease, and that the oil price has a significant effect on the real exchange rate movements. At the same time, Egert and Leonard (2011) examined the Kazakh economy for the presence of Dutch disease and tested for the B-S effect during the years 1996 through 2005 when oil prices had been rising. Their conclusion was that the non-oil tradable sector was unaffected by the increase in oil revenues and that the appreciation of the currency was mostly due to the change in the nominal rate instead of an increase in the price level.

Amin and El-Sakka (2016) found a long-run relationship between oil prices, GDP and real exchange rates of dollar-pegged GCC countries and noted that there is causality going from oil prices to the exchange rate, but that the adjustment of the exchange rate to the equilibrium is very slow.

Due to limitations in publicly available data, most studies that analyse oil-exporting countries, focus on the effects of changes in the oil price rather than on country-level changes in productivity. For example, Habib and Kalamova (2007) found no relationship between oil prices and real exchange rates in Norway and Saudi Arabia, but established a positive relationship in Russia. Aziz and Abu Bakar (2009), however, found no long-run relationship for net oil exporters – Canada, Denmark and Malaysia, unlike net oil importers, which in their sample appeared to have negative relationship between the oil price and their currency values. More discussions on the relationship between the oil price and the real exchange rates in oil-rich countries can be found in Rickne (2009) and Frankel (2017).

Although the literature on B-S hypothesis is extensive and spans several decades and econometric approaches, we hope to fill three main gaps: 1) Analysing the effect with data on country-level productivity (rather than revenue); 2) Using a more diverse set of oil-exporting countries, and 3) Implementing panel estimation methods that require large-N and large-T data structure. The previous literature points to the second point as a possible extension of current work, and the first and third points are feasible only with a dataset like ours. We now continue with a discussion of our dataset and descriptive statistics.

3.4 Data.

3.4.1 Sources and format

The main sources of data are Penn World Table version 7.1, Wood Mackenzie's (WM) Global Economic Model (GEM), and BP's Statistical Review of World Energy. Table 3.2 below lists all the key variables and their sources.

Variable	Abbreviation	Description	Source
Real exchange rate	*rer*	Real exchange rate (local currency units per I$)	PWT 7.1
Oil rents per capita	*oilrents_pc*	Total oil rents (constant million 2005 US$) divided by population	Wood Mackenzie & BP
Real GDP per capita	*rgdpch*	PPP converted GDP per capita chain series (2005 I$)	PWT 7.1
Brent price	*brent*	Brent oil price (2005 US$)	Thomson Reuters Datastream
Openness to trade	*openc*	Openness at current prices (%)[6]	PWT 7.1

Table 3.1. *Key variables and sources*

The dataset has a panel format and covers 42 countries over up to 45 years – 1965 through 2009. However, the panel is unbalanced with partial gaps in most countries' time series. The shortest time series available are for India, Azerbaijan, and China with 11, 12 and 13 years of data. On average, the dataset has 28 years of data for each of the 42 countries for a total of 1190 observations of oil rents. Table 3.3 provides summary statistics for key variables and their natural logarithms, where appropriate. Table 3.4 below outlines the number of years available for each country.

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Oil rents per capita	1190	1426	3374	-20389	39331
ln(oil rents per capita)	1190	5.00	3.29	-9.92	10.58
Real exchange rate	1190	1.75	1.01	0.21	15.00
ln(real exchange rate)	1190	0.44	0.49	-1.54	2.71
Real GDP per capita	1190	14270	16465	612	118771
ln(real GDP per capita)	1190	8.97	1.12	6.42	11.68
Brent	45	38.43	19.77	15.07	91.12
ln(brent)	45	3.53	0.48	2.71	4.51
Openness to trade	1190	73.98	40.07	14.68	354.11

Table 3.3 Coverage by variable (N + T dimension)

[6] Defined as the sum of exports and imports as a fraction of GDP

Countries / Variables	Country code	Number of years	OPEC	D10[7]
Algeria	DZA	45	✓	✓
Angola	AGO	24	✓	✓
Argentina	ARG	18	-	-
Australia	AUS	40	-	-
Azerbaijan	AZE	12	-	✓
Brazil	BRA	17	-	-
Brunei	BRN	40	-	✓
Canada	CAN	29	-	-
China	CHN	13	-	-
Colombia	COL	24	-	-
Congo, Republic of	COG	36	-	✓
Denmark	DNK	37	-	-
Ecuador	ECU	13	✓	✓
Egypt	EGY	45	-	✓
Equatorial Guinea	GNQ	18	-	✓
Gabon	GAB	45	-	✓
India	IND	11	-	-
Indonesia	IDN	42	-	-
Iraq	IRQ	40	✓	✓
Italy	ITA	45	-	-
Kazakhstan	KAZ	15	-	✓
Libya	LBY	37	✓	✓
Malaysia	MYS	37	-	✓
Mexico	MEX	16	-	-
Nigeria	NGA	36	✓	✓
Norway	NOR	32	-	✓
Oman	OMN	36	-	✓
Peru	PER	30	-	-
Qatar	QAT	24	✓	✓
Romania	ROM	21	-	-
Russia	RUS	25	-	✓
Saudi Arabia	SAU	40	✓	✓
Sudan	SDN	15	-	✓
Syria	SYR	36	-	✓
Thailand	THA	25	-	-
Trinidad &Tobago	TTO	35	-	✓
Tunisia	TUN	43	-	-
United Arab Emirates	ARE	24	✓	✓
United Kingdom	GBR	41	-	-
Venezuela	VEN	21	✓	✓
Vietnam	VNM	20	-	✓
Yemen	YEM	24	-	✓

Table 3.4 Coverage by country (time dimension) and subsample composition

[7] Countries, in which oil rents exceed 10% of GDP in 2008

In addition to the variables that directly measure countries' productivity levels – oil rents per capita and real GDP per capita – we have also included the Brent oil price as well as openness to trade at current prices. Both variables have been widely used as controls in the Balassa-Samuelson literature. Oil price is included to control for the direct impact of its change on the real exchange rate of the oil-exporting countries, which has been shown to be the case for oil exporters (see Chinn (2000), Egert et al. (2006), and Korhonen and Juurikkala (2009)). Openness to trade has also been shown to affect the real exchange rate directly, however, unlike the oil price, which has a negative relationship with the real exchange rate, there is some controversy regarding the sign of the relationship (see Egert et al. (2006) for discussion). Most studies, however, find that trade liberalisation results in currency depreciation (Kim and Korhonen (2005), Njindan Iyke (2017)), so the sign of the relationship is expected to be positive.

3.4.2 Construction of oil rents per capita and its natural log

As an intermediate step to calculating total oil rents, we calculate a cost ratio. This interim variable is the ratio of total costs and gross revenue from Wood Mackenzie's GEM. The former consists of capital and operating costs, which are summed to get total costs. Due to the nature and coverage of GEM data (further explained in section 3.4.3 below), revenues from Wood Mackenzie are not used directly in our estimations. Instead, we calculate oil revenues using BP production figures and Brent price series obtained from Datastream. These steps are summarised in table 3.5:

$$Total\ costs = Capex\ (WM) + Opex\ (WM)$$

$$Cost\ ratio = \frac{Total\ costs\ (WM)}{Gross\ revenue\ (WM)}$$

$$Oil\ revenue == Oil\ production\ (BP) \times Brent\ price$$

$$Oil\ rents = Oil\ revenue \times (1 - Cost\ ratio)$$

Table 3.5. Construction of oil rents

3.4.3 Data treatment and limitations

All key variables with the exception of $openc$—expressed as a %—are used in natural logarithm form.[8] One important limitation of Wood Mackenzie's GEM data was its coverage. The database was structured based on concessions and exploration license areas. To get an idea of the whole country's petroleum industry, we aggregated the granular data points. In some cases, however, Wood Mackenzie's coverage of the country's production was limited to certain areas only. This posed a problem for the oil rents variable, since the underlying assumption of our calculation shown in section 3.4.2 is that the cost ratio is applicable to the whole country. This may not hold if the GEM database's coverage of the country is quite limited. To resolve this, we impose a restriction that an oil rents observation is used only if WM's field-by-field cost estimations cover at least 10% of the country's production as listed by BP. This procedure discards observations where the WM data is less likely to be representative of the country as a whole. As far as we know, the gaps in Wood Mackenzie's data are basically random[9] with respect to the relationship that we are esimating, as they have arisen from the histories of the various data sets which were collected or acquired by Wood Mackenzie around the world at different times; the gaps are largest in the first years in which WM has data on a country, but this is not necessarily when the country began production. Thus dropping these observations should not bias our results. Figures

[8] In a small number of year and country combinations, per capita oil rents were observed to be negative. These tend to occur in countries with relatively small economies and in years just after a discovery when substantial upstream investment is taking place within the petroleum industry such that total costs exceed gross revenue (examples include Australia, in 1970-71, Brunei in 1970-1973, etc). The negative numbers posed a data treatment issue while transforming the variable with the natural log function. To preserve symmetry during the transformation, we used the negative log of the absolute values of negative entries. There are of course a variety of approaches when dealing with negative observations in the context of a logarthmic transformations (dropping the observations, Winsorizing, adding a constant, etc.) but: (i) these approaches generally did not preserve symmetry, and (ii) the choice of transformation does not appear to affect any of the main qualitative findings anyhow.

[9] The 10% threshold affects 8 out of 42 countries (Argentina, Brazil, Canada, China, Columbia, Ecuador, India and Romania) in at least one year, and 71 out of 1261 observations overall. The years affected span the whole range from 1960s to the 1990s, with the average year of a dropped observation being 1982 relative to the whole sample average of 1993.

3.1 and 3.2 below facilitate visualisation of the resulting series for all the countries in the dataset in two selected years – 2000 and 2009, respectively.

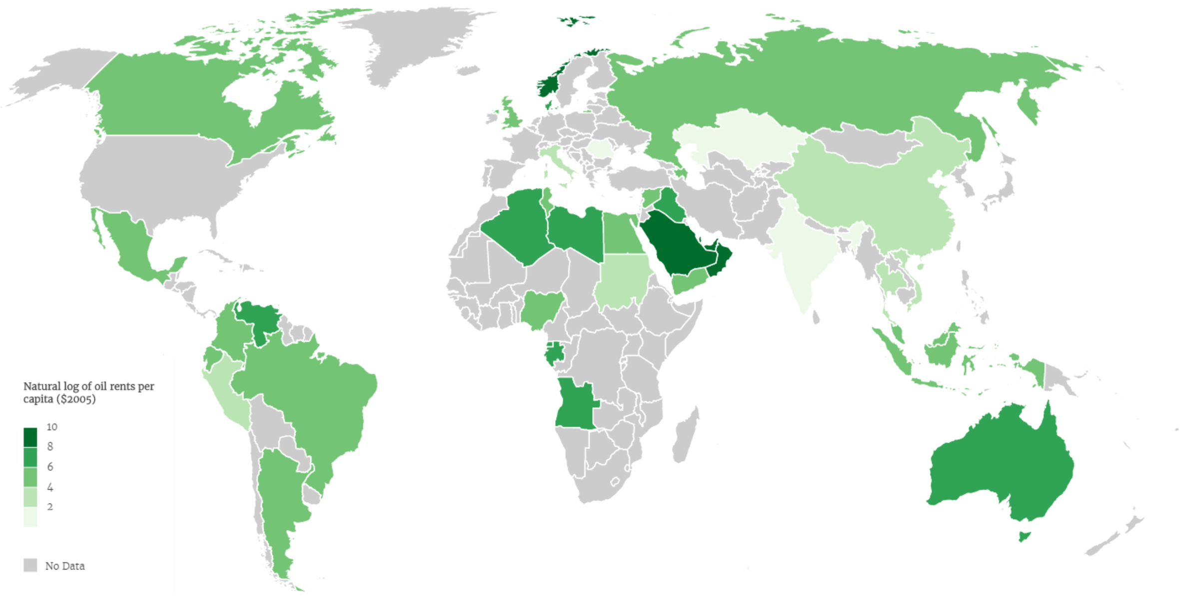

Figure 3.1. Natural logarithm of oil rents per capita in 2000

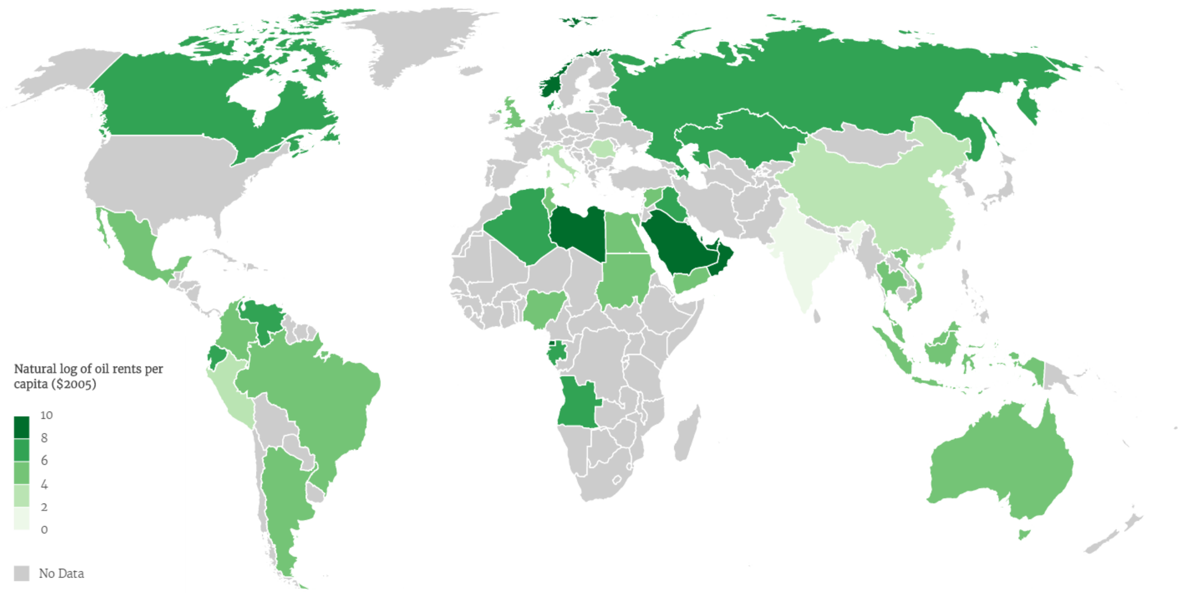

Figure 3.2. Natural logarithm of oil rents per capita in 2009

3.4.4 Country coverage and subsamples

Since we expect to observe a stronger evidence for our hypothesis in oil-dependent countries, we focus on subsamples of countries: OPEC countries and those in which oil rents exceed 10% of GDP in 2008 – referred to as “D10” countries (these categories are obviously not exclusive). Table 3.4 above shows which countries OPEC and D10 categories consist of and table 3.6 below summarises the key variables for each subsample.

Oil rents per capita in OPEC and D10 countries are considerably higher than in the rest of the world – the mean oil rents (in 2005 US$) per capita for OPEC and D10 countries are about $2800 and $2200 per capita, respectively, whereas that for the rest of the countries is $150 per capita. Note here that OPEC is a subset of D10 countries and that not all OPEC countries are covered by our dataset. It should be noted that Indonesia wasn’t included into OPEC countries because it was often a marginal member due to relatively low exports – it originally joined in 1962, but left in 2009 after being a net importer for some years, then re-joined in 2016 and finally suspended its membership by the end of the year.

Variable	OPEC			
	Mean	Standard Deviation	Minimum	Maximum
Oil rents per capita	2821	4252	-114	32066
ln(oil rents per capita)	6.88	1.84	-4.74	10.38
Real exchange rate	1.67	1.11	0.21	7.76
ln(real exchange rate)	0.35	0.61	-1.54	2.05
Real GDP per capita	16161	21983	976	118771
ln(real GDP per capita)	8.90	1.24	6.88	11.68
Openness to trade	85.59	48.72	23.61	354.11
N / n	10 / 275			
	D10			
	Mean	Standard Deviation	Minimum	Maximum
Oil rents per capita	2206	4090	-20389	39331
ln(oil rents per capita)	5.94	3.26	-9.92	10.58
Real exchange rate	1.83	1.16	0.21	15.00
ln(real exchange rate)	0.46	0.53	-1.54	2.71
Real GDP per capita	14467	18992	612	118771
ln(real GDP per capita)	8.86	1.20	6.42	11.68
Openness to trade	87.30	42.41	14.77	354.11
N / n	26 / 738			
	World (excl. D10)			
	Mean	Standard Deviation	Minimum	Maximum
Oil rents per capita	151	233	-237	1534
ln(oil rents per capita)	3.47	2.73	-5.47	7.34
Real exchange rate	1.62	0.69	0.64	4.88
ln(real exchange rate)	0.40	0.41	-0.45	1.58
Real GDP per capita	13948	11185	712	40820
ln(real GDP per capita)	9.15	0.96	6.57	10.62
Openness to trade	52.24	23.02	14.68	151.71
N / n	16 / 452			
	World			
	Mean	Standard Deviation	Minimum	Maximum
Oil rents per capita	1426	3374	-20389	39331
ln(oil rents per capita)	5.00	3.29	-9.92	10.58
Real exchange rate	1.75	1.01	0.21	15.00
ln(real exchange rate)	0.44	0.49	-1.54	2.71
Real GDP per capita	14270	16465	612	118771
ln(real GDP per capita)	8.97	1.12	6.42	11.68
Openness to trade	73.98	40.07	14.68	354.11
N / n	42 / 1190			

Table 3.6. Descriptive statistics by subsample

To get a better understanding of how the key variables behave, we refer to the following figures.

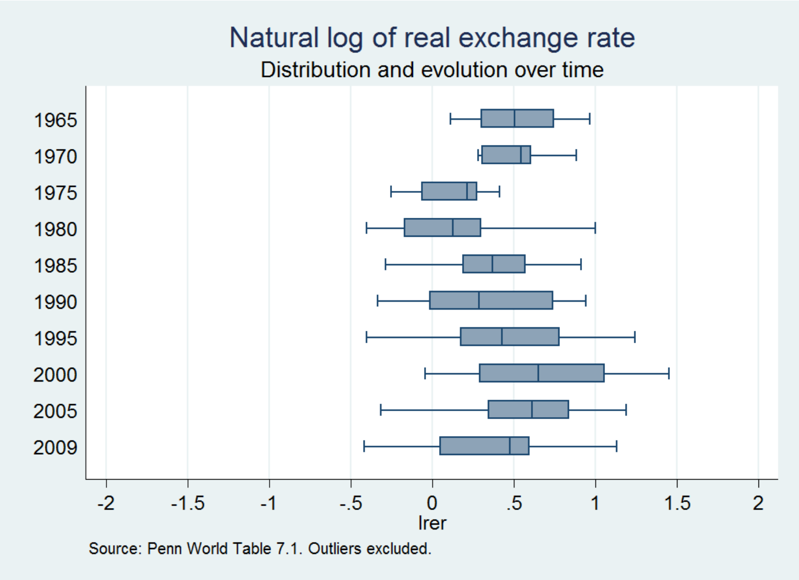

Figure 3.3. Natural log of real exchange rate across time

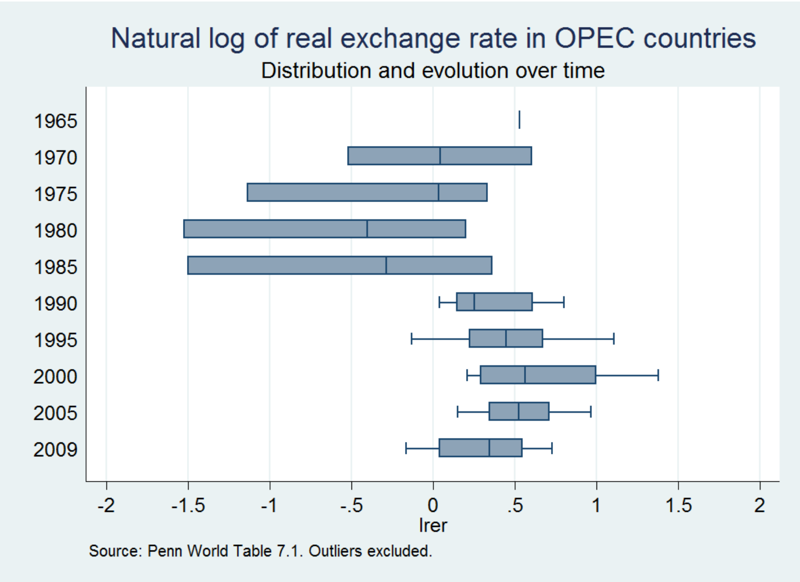

Figure 3.4. Natural log of real exchange rate across time in OPEC countries

Figure 3.3 and 3.4 show the distribution and evolution over time of the real exchange rate in OPEC countries and the rest of the world. With the exception of the early years in the dataset, OPEC countries' real exchange rates behave similarly to the rest of the world.

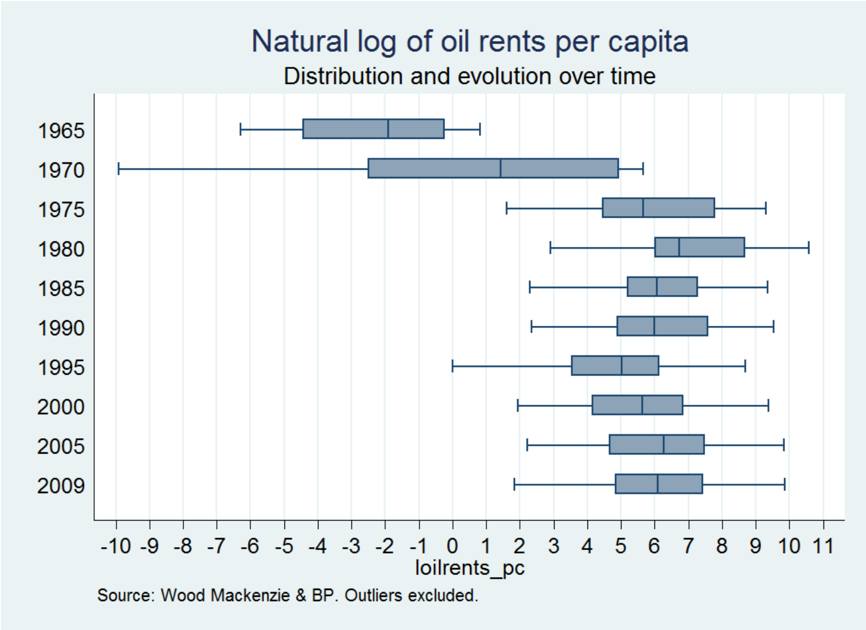

Figure 3.5. Natural log of oil rents per capita

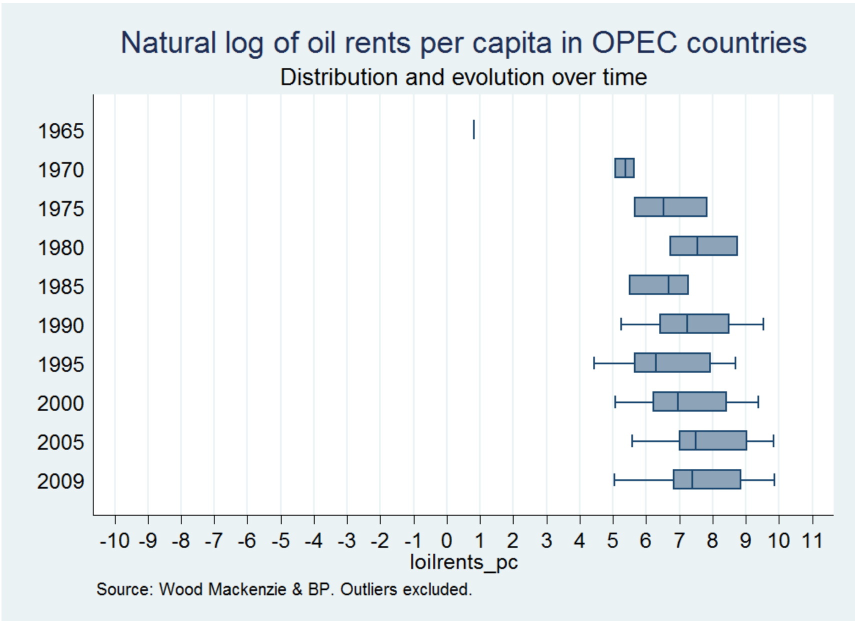

Figure 3.6. Natural log of oil rents per capita in OPEC countries

In figures 3.5 and 3.6, we turn to our main explanatory variable, oil rents per capita. Unsurprisingly and as previously observed, OPEC countries have higher per capita oil rents than the rest of the world.

Figures 3.7 and 3.8 below (GDP figures) show that although mean GDP for OPEC countries is very similar to the global average (see table 3.6), the series behaves differently across time. For instance, global GDP appears to have a positive time trend, whereas this is much less pronounced in OPEC countries. This is partially due to the fluctuations in oil prices or oil production in some years, since oil rents constitute a relatively large portion of these countries' GDPs.

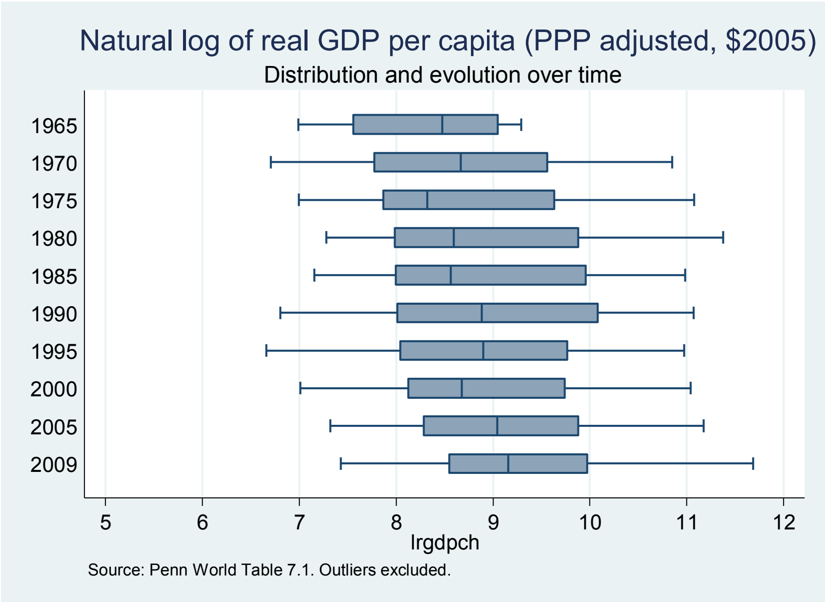

Figure 3.7. Natural log of real GDP per capita over time

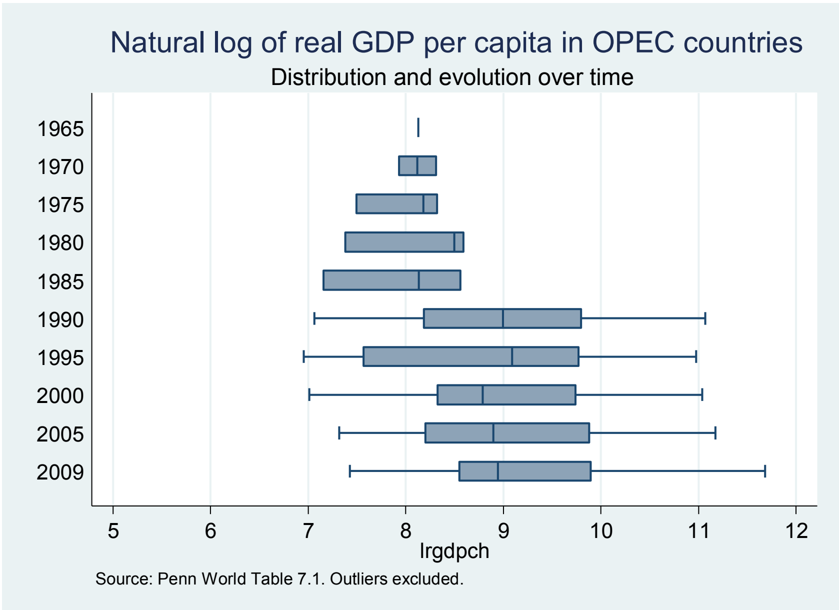

Figure 3.8. Natural log of real GDP per capita over time in OPEC countries

Finally, figures 3.9 and 3.10 below combine the key variables in question. Casual inspection reveals a negative relationship between real exchange rate and oil rents per capita. This is along the lines of what theory predicts and will be discussed in detail in the rest of the chapter. Interestingly, the OPEC subsample has a clearer and more negative relationship based on figure 13.10. This provides preliminary evidence for B-S hypothesis and will be investigated further.

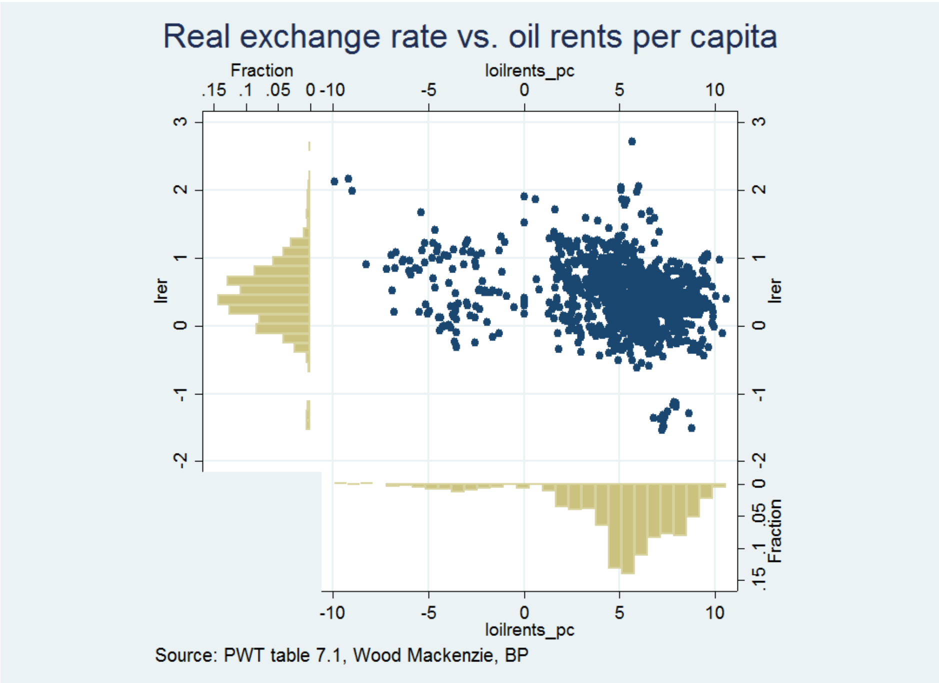

Figure 3.9. Real exchange rate versus oil rents per capita

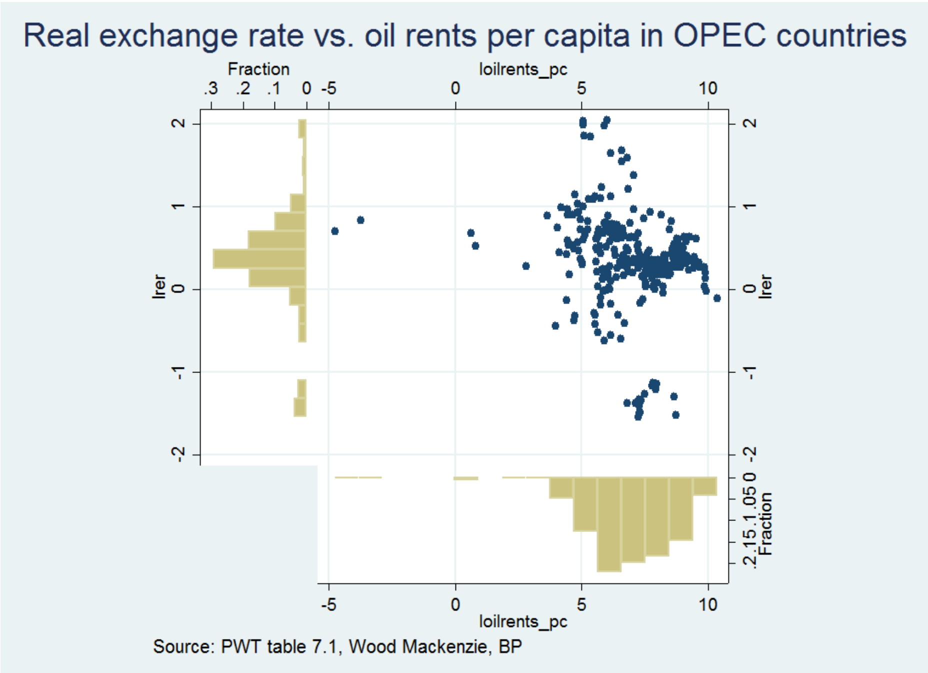

Figure 3.10. Real exchange rate versus oil rents in OPEC countries

3.4.5 Individual country statistics and plots

This section aims to provide further clarity on individual countries in our dataset by summarising and graphing key variables. Generally, we observe a negative relationship between the logarithm of real exchange rates, $lrer$, and the main explanatory variable, which is the logarithm of per capita oil rents, $loilrents\_pc$. The simple correlation coefficient between the two variables using the whole dataset is -0.26.[10] This coefficient is -0.32 when the sample is restricted to D10 countries. It is, therefore, not surprising that we observe this relationship in most of the countries in the dataset. Angola and Norway are shown below as examples.

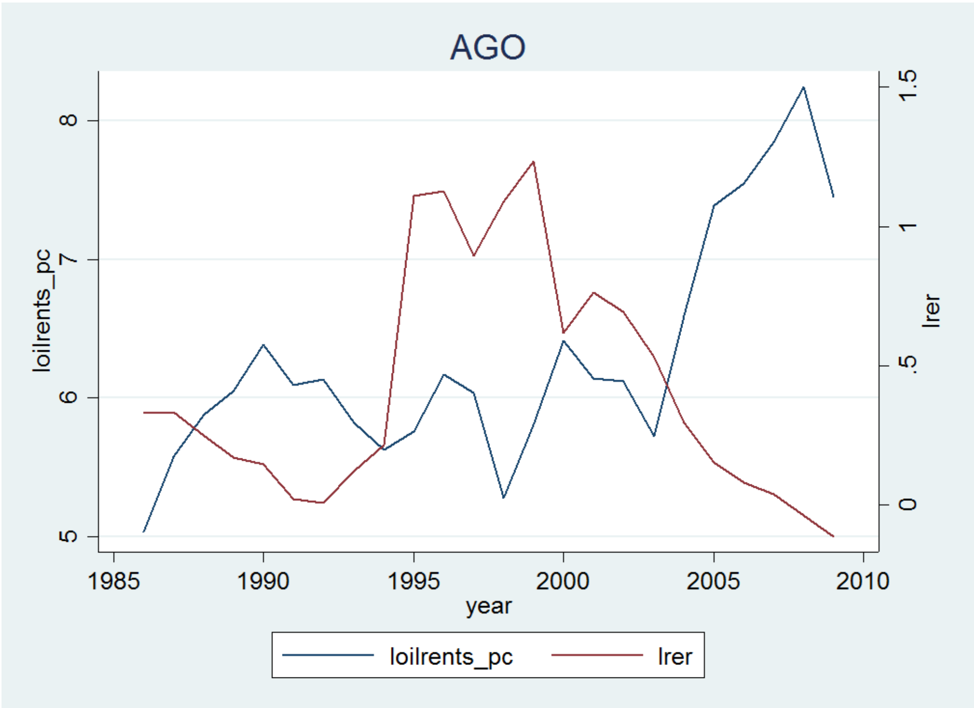

Figure 3.11. Real exchange rates and oil rents per capita in Angola

[10] Recall that the exchange rate here is defined as national currency per US dollar, so that an increase in the variable represents a depreciation. Thus the -0.26 correlation implies that a rise in oil rents is associated with an appreciation (i.e. a decrease in *lrer*).

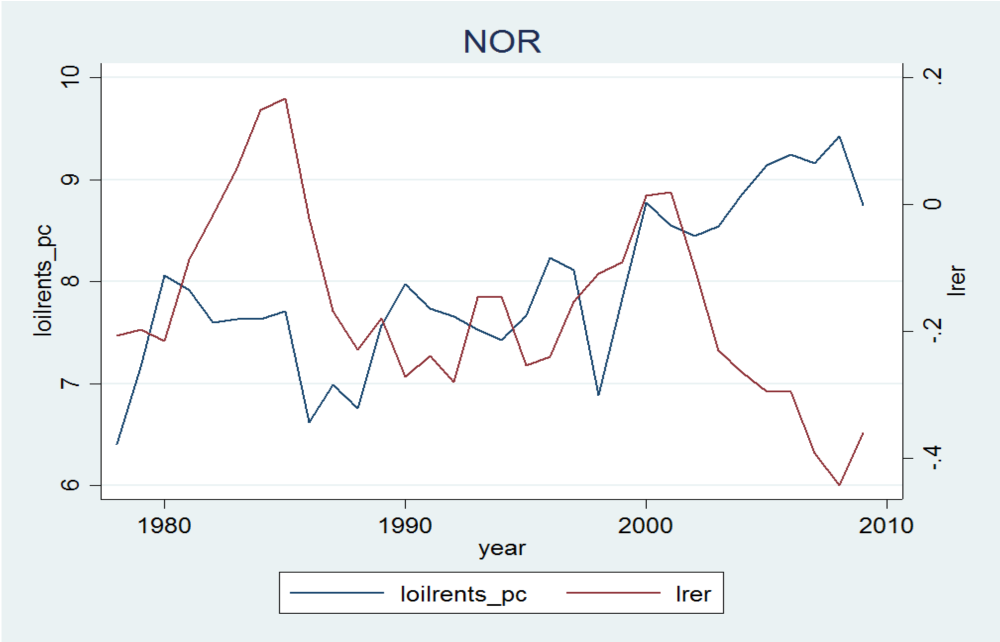

Figure 3.12. Real exchange rates and oil rents per capita in Norway

Unsurprisingly, not all countries follow this pattern. An example is Australia, where the real exchange rate fluctuates while per capita oil rents have been fairly stable. The Australian real exchange rate is mainly driven by other factors – the mean oil rents in Australia are 1.3% of GDP and they have never exceeded 4% of GDP.

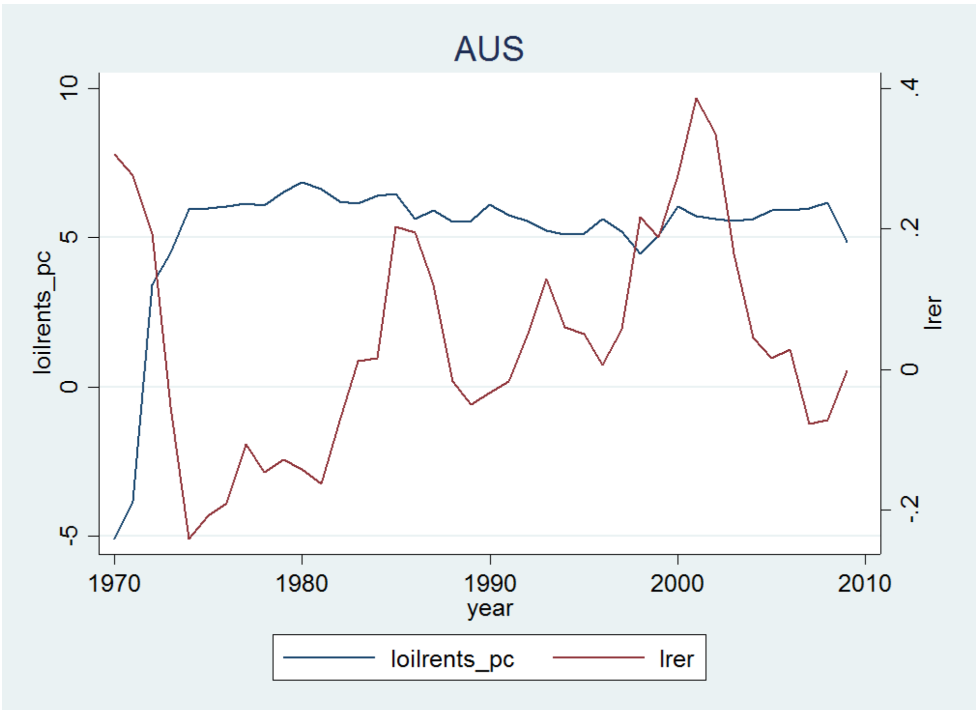

Figure 3.13. Real exchange rates versus oil rents per capita in Australia

3.5 Unit root tests.

The first step of the analysis is to test whether the variables are stationary or non-stationary, since this will determine the appropriate estimation technique to be used later.

One of the early panel unit root tests is the Levin and Lin test, which was first introduced in 1992. Nowadays, the most commonly used tests in applied research are LLC (Levin, Lin and Chu (2002)), IPS (Im, Pesaran and Shin (2003)), Hadri (2000) and the Fisher-type tests (1932). Preference is generally given to tests that can be applied to unbalanced panels, that is tests that do not assume a common unit root process (like LLC and Hadri) or require the same number of observations in all the panels (like IPS). To overcome these problems we use the Fisher-type test, proposed by Maddala and Wu (1999), which carries out individual independent unit root tests and then combines the p-values. The test doesn't restrict panels to have the same number of observations or the same number of lags. Another useful property of the test is that it can combine significance levels from different individual unit root tests. In order to potentially improve on the size and power properties, another option would be to restrict the samples to have the same number of observations to perform other panel unit root tests. However Madala and Wu (1999) showed that the Fisher test performs better in the Monte-Carlo studies compared to LLC or IPS, so given the trade-off introduced by the loss of some observations, the Fisher-type test would seem to be the best choice here. Also, Pesaran (2012) notes that there is no theoretical basis for the homogeneous AR structure in the context of testing the PPP hypothesis, which makes LLC and Hadri tests even less attractive for this analysis. Davidson and Mackinon (2004) compared the ADF to the PP test and found the ADF test to exhibit better power properties in finite samples.

The nature of the data in principle would assume N (number of panel units) is fixed and T (number of time periods) tends to infinity, which is reasonable for some panels. However, we have an unbalanced panel and N is larger than an average number of observations per country.

Table 3.7 below reports the results of the Fisher-type test that utilises individual ADF regressions. The number of lags varies from zero to two, and all tests are performed with and without a trend. The null hypothesis states that all the panels contain unit root. The default significance level is chosen to be 5%.

Variable	Trend	Number of lags	Test statistic	Probability	Stationary or not
lrer	No	0	82.53	0.525	All panels are non-stationary
	No	1	139.81	0.000	At least some panels are stationary
	No	2	122.10	0.004	At least some panels are stationary
	Yes	0	84.29	0.471	All panels are non-stationary
	Yes	1	181.74	0.000	At least some panels are stationary
	Yes	2	135.01	0.000	At least some panels are stationary
loilrents_pc	No	0	229.41	0.000	At least some panels are stationary
	No	1	206.76	0.000	At least some panels are stationary
	No	2	135.83	0.000	At least some panels are stationary
	Yes	0	189.61	0.000	At least some panels are stationary
	Yes	1	201.63	0.000	At least some panels are stationary
	Yes	2	129.34	0.001	At least some panels are stationary
lrgdpch	No	0	67.58	0.905	All panels are non-stationary
	No	1	80.99	0.573	All panels are non-stationary
	No	2	81.70	0.551	All panels are non-stationary
	Yes	0	70.87	0.846	All panels are non-stationary
	Yes	1	130.73	0.001	At least some panels are stationary
	Yes	2	71.73	0.828	All panels are non-stationary
openc	No	0	159.09	0.000	At least some panels are stationary
	No	1	132.40	0.001	At least some panels are stationary
	No	2	98.88	0.128	All panels are non-stationary
	Yes	0	132.85	0.001	At least some panels are stationary
	Yes	1	124.98	0.002	At least some panels are stationary
	Yes	2	64.61	0.943	All panels are non-stationary

Table 3.7. Panel unit root tests.

It can be seen from table 3.7 that $lrer$ appears to be non-stationary if no lags are included, but when at least one lag is included we reject the null of non-stationarity of all the panels. $openc$ exhibits somewhat opposite behaviour - when two lags are included we can't reject the null of non-stationarity of all the panels. The Fisher test strongly suggests that $lrgdpch$ is non-stationary – the results hold for five out of six specifications. However, inclusion of a trend combined with 1 lag rejects the null. For $loilrents\_pc$ the null is always rejected.

The Fisher-type tests have a significant drawback – the null hypothesis of non-stationarity implies all panels are non-stationary against the alternative that some are stationary which isn't a particularly interesting or useful question to answer. Combined with low power properties when the number of observations is small, the test is informative, but ultimately looking at the individual unit root tests can be more practical as it is unlikely that variables in *all* the panels will exhibit the same order of integration, especially given the number of observation differs by country. The individual ADF test results are presented in tables A.1-A.4 of the Appendix and summarised results are presented in table 3.8.

Variable	**Trend**	**Number of lags**	**Number of countries (out of 42) for which the null hypothesis[11] is rejected at the 5% s.l.**	**Fraction of countries for which the null hypothesis is not rejected**
loilrents_pc	No	0	10	76%
loilrents_pc	No	1	6	86%
loilrents_pc	No	2	5	88%
loilrents_pc	Yes	0	8	81%
loilrents_pc	Yes	1	8	81%
loilrents_pc	Yes	2	6	86%
lrer	No	0	3	93%
lrer	No	1	8	81%
lrer	No	2	5	88%
lrer	Yes	0	4	91%
lrer	Yes	1	7	83%
lrer	Yes	2	6	86%
lrgdpch	No	0	2	95%
lrgdpch	No	1	2	95%
lrgdpch	No	2	4	91%
lrgdpch	Yes	0	1	98%
lrgdpch	Yes	1	5	88%
lrgdpch	Yes	2	2	95%
openc	No	0	4	91%
openc	No	1	5	88%
openc	No	2	2	95%
openc	Yes	0	5	88%
openc	Yes	1	5	88%
openc	Yes	2	0	100%

Table 3.8. Country-by-country ADF tests summarised.

$lbrent$ was tested using Elliot, Rothenberg and Stock (1996) (ERS) DF-GLS test, which is a more powerful modification of the ADF test. The lag length is chosen according to MAIC, the modified Akaike information criterion (Ng and Perron (2001)), and set to 1, however the null hypothesis is not rejected at the 10 % significance level for up to 9 lags, which confirms that $lbrent$ is non-stationary around a linear trend. 45

[11] The null hypothesis is that the unit root is present

observations are utilised, which corresponds to the maximum number of observations in the dataset.

Variable	Trend	Number of lags	Test statistic	Rejection	Stationary or not
lbrent	No	1	-1.41	Do not reject at 10% s.l.	Series is non-stationary
lbrent	Yes	1	-1.82	Do not reject at 10% s.l.	Series is non-stationary

Table 3.9. Log of Brent oil price unit root test results.

Table 3.8 suggests that for all five variables most panels are non-stationary – according to a combination of Schwert (1989) criterion, which determines the maximum lag length, and MAIC the optimal lag length varies from 1 to 2 lags for all tests, so for at least 80% of all panels the null of non-stationarity cannot be rejected. We will proceed the analysis treating all variables as non-stationary.

3.6 Methodology and estimations.

3.6.1 Testing for cointegration

Having just established the non-stationarity properties of the variables, the next step is to test for the existence of the cointegrating relationship between *lrer* and *oilrentc_pc* and potentially other determinants of the real exchange rate. If the variables prove to have a long-run relationship, then the dynamic ordinary least squares estimator (DOLS) introduced by Stock and Watson (1993), the mean-group estimator (MG) of Pesaran and Smith (1995) and the pooled mean-group estimator (PMG) of Pesaran et al. (1999) will be used to evaluate the magnitude of the coefficients in the long-run equation. The short-run dynamics will also be discussed. The DOLS estimator has been frequently used in the literature in the context of the B-S hypothesis, for example MacDonald and Ricci (2007) and Chong et al. (2012), but the MG and PMG estimators are not yet prevalent in the B-S literature, despite some analyses along these lines, such as Camarero (2008).

OLS estimation of the cointegrated non-stationary variables also produces consistent estimates, but they generally do not follow a Gaussian distribution, so the conventional test statistics are meaningless. Stock and Watson's (1993) DOLS was suggested as a solution to this problem. The estimator is asymptotically efficient and normally distributed - this is achieved by the inclusion of the leads and lags of the differenced explanatory variables, which orthogonalizes the error term with respect to the innovations in the regressors. This has the added advantage of eliminating potential endogeneity between the error term and the stationary component of the non-stationary variables. The asymptotically valid standard errors can be computed using a HAC estimator; Newey-West standard errors will be used for these purposes. According to Kao and Chiang (2001) DOLS outperforms panel OLS and fully modified least squares estimator (FMOLS) by having smaller bias and smaller finite sample size distortions.

The DOLS equation has the following form:

$$lrer_{it} = X_{it}\,\beta + \alpha_i + \sum_{k=-a}^{b} \Delta X_{it+k}\theta + \epsilon_{it} \tag{3.1}$$

where $lrer_{it}$ is the real exchange rate of country i at time t measured in natural logarithm, X_{it} is the vector of explanatory variables, β is the vector of the long-run DOLS coefficients, α_i are the country fixed effects, θ is a vector of the coefficients on the lags and leads of the first-differenced explanatory variables, and ε_{it} denotes the error term. Maximum lag and lead lengths are shown by a and b, respectively.

To test for the presence of cointegration in the context of panel data, Pedroni (1999, 2004) suggested seven test-statistics in the Engle-Granger tradition. The null hypothesis of the tests is no cointegration. Four statistics are panel, three are group. Panel statistics assume homogeneity of the panels and pool the data across the within dimension, constraining the coefficients to be the same. Group statistics allow for heterogeneity of the panels and calculate averages for the statistics from individual time-series estimations. The latter are more relevant for our estimation, because it would be reasonable to assume that coefficients could vary across countries. The residuals for the Pedroni test are obtained from the long-run DOLS equation that has the following form:

$$lrer_{it} = loilrents\_pc_{it}\,\beta + \alpha_i + \sum_{k=-1}^{1} \Delta\, loilrents\_pc_{it+k}\theta + \epsilon_{it} \tag{3.2}$$

The results for the Pedroni tests are presented in table 3.10. In all specifications for this test and for the subsequent DOLS estimation, one lead and one lag are used in order to avoid constraining the number of observations, particularly in the case of countries with reduced number of years - a larger lag length would only be feasible if some countries are dropped. Furthermore, our main findings are robust to an increased number of leads and lags in constrained datasets. The table indicates that four out of seven tests reject the null of no-cointegration at the 10% significance level for the whole sample and even more strongly for the D10 countries. However, there seems to be no cointegrating relationship found in countries – none of the seven statistics is significantly different from zero (the test statistic has standard normal distribution under the null). The result is somewhat unexpected; it suggests that the productivity bias hypothesis doesn't explain movements in the real exchange rate through the productivity of the oil sector in the countries where the oil sector accounts for a large fraction of tradable exports.

Test statistic		(1)	(2)	(3)
Panel	v-statistic	2.823***	1.906**	0.463
	rho-statistic	-2.021**	-2.109**	-0.963
	t-statistic	-1.611	-2.203**	-1.370
	ADF-statistic	-1.016	-1.508	-0.785
Group	rho-statistic	0.136	-0.822	0.619
	t-statistic	-1.876*	-3.180***	-0.532
	ADF-statistic	-1.915*	-0.530	-0.045
Subsample		World	D10	OPEC
N		42	26	10
Lags		1	1	1

Table 3.10 Pedroni (1999, 2004) cointegration test results.

Note: *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

As the Pedroni test results don't appear to be conclusive for the whole sample and D10 countries, we conduct Westerlund (2007) cointegration tests, which unlike the Pedroni tests can be made robust to cross-sectional dependence. The tests are based on estimating an ECM-type equation (equation 3.3) and testing whether the panels are error-correcting. The original test has four test statistics – all four share a common null hypothesis, but they have different alternative hypotheses. The tests are: Gt, Ga, Pt and

Pa. Gt and Ga test the null of no cointegration ($\varphi_i = 0\ for\ all\ i$) against the alternative that at least one panel contains cointegration ($\varphi_i < 0\ for\ at\ least\ one\ i$). Pt and Pa also have a null hypothesis of no cointegration ($\varphi_i = 0\ for\ all\ i$), but the alternative is stated as all panels exhibit cointegration ($\varphi_i < 0\ for\ all\ i$). We will focus on Pt and Pa because some panels have very small numbers of observations and Gt and Ga are more likely to suffer from low power, which is exacerbated by the inclusion of the lags and leads into the ECM used to estimate the residuals – even fewer degrees of freedom are left for the test. The lag length is again set to one to preserve the sample, however, if the sample is restricted to countries with larger number of observations and the optimal leg length is selected according to the Akaike information criterion (AIC) the results remain unchanged.

$$\Delta lrer_{it} = \varphi_i(lrer_{it-1} - \beta_i loilrents\_pc_{it-1}) + \alpha_i + \sum_{k=1}^{p} \Delta\, lrer_{it-p}\delta_i + \sum_{k=0}^{p} \Delta\, loilrents\_pc_{it-k}\theta_i + v_{it} \quad (3.3)$$

where φ_i is a speed of adjustment coefficient, δ_i is a (px1) vector of coefficients on the lagged first-differenced dependent variable and θ_i is a ((p+1)x1) vector of coefficients on the lagged first-differenced regressor, v_{it} denotes the error term.

The results of the Westerlund cointegration test are presented in table 3.11. It can be seen in columns (1) and (2) that for both the whole sample and D10 countries, the Pt statistic suggests that the null of no cointegration should be rejected, whereas Pa cannot reject the null at any conventional significance level. Column (3), similarly to Pedroni test results, provides strong evidence against any cointegrating relationship in OPEC countries. Columns (4) and (5) retest the hypothesis for the whole world and D10 countries with OPEC countries being excluded from the samples and the results strongly indicate cointegration for the remaining countries – all four test statistics are rejected at the 1% significance level.

Test statistic		(1)	(2)	(3)	(4)	(5)
	z-value	-2.518***	-1.893**	-0.532	-3.732***	-2.915***
Pt	p-value	0.006	0.029	0.297	0.000	0.002
	z-value	-1.183	-0.568	0.518	-3.843***	-2.758***
Pa	p-value	0.118	0.285	0.698	0.000	0.003
Subsample		World	D10	OPEC	World-OPEC	D10-OPEC
N		42	26	10	32	16
Lags		1	1	1	1	1

Table 3.11. Westerlund (2007) cointegration tests.

Note: *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

Cointegration tests have also been performed using other explanatory variables - oil price and real per capita GDP instead of the oil sector productivity. The results suggest that the real exchange rate both for the whole sample and for D10 countries (excluding OPEC countries) is cointegrated with the oil price, while no cointegration is found in OPEC countries. Real GDP per capita doesn't exhibit any long-run relationship with the real exchange rate in the whole sample or in D10 countries, but there is weak evidence of potential relationship in OPEC countries (see table A.5 of the Appendix for the results).

Summing up, the cointegration tests carried out have strongly suggested that there is a long-run relationship between the real exchange rate and the productivity of the oil sector in most oil-exporting countries in our sample, but this relationship does not exist in OPEC countries. However, OPEC countries are kept in the sample for the future analyses, so the results that follow below have to be interpreted with this caveat. Possible explanations for the failure of the B-S hypothesis in OPEC countries are discussed in section 3.6.4.

3.6.2 Dynamic OLS Results

As previously discussed, DOLS estimations are used as a robust alternative to Panel OLS. The number of observations for some panel units are limited and pooling the data can improve the power of our results. The main interest lies in the long-run relationship between the real exchange rate and the main explanatory variable, oil rents per capita. A panel cointegration model using DOLS is implemented to treat the non-stationarity of our variables appropriately. As noted earlier, this approach has been used extensively within the B-S literature. Similarly to section 3.6.1, one lag and one lead are used such that $a = b = 1$, and our main findings are robust to an increased number of leads and lags. Time fixed effects are excluded in our model, as these were observed to have little effect on our results. Furthermore, unlike in Gubler and Sax (2011), where time fixed effects were critical, our dependent variable is calculated relative to the US dollar.

Having established the existence of panel cointegration in non-OPEC countries in section 3.6.1, we interpret β in equation 3.1 as the long-run coefficient. In addition to estimating this long-run relationship, we include an error correction specification to capture the short-run dynamic adjustment of the real exchange rate towards equilibrium. The error correction model (ECM) has the following form:

$$\Delta lrer_{it} = \gamma + \varphi gap_{it-1} + \sum_{j=1} \Delta lrer_{it-j}\phi_j + \sum_{j=0}^{1} \Delta X_{it-j}\omega_j + \nu_{it} \tag{3.4}$$

where

$$gap_{it} = lrer_{it} - X_{it}\,\beta - \alpha_i - \sum_{k=-1}^{1} \Delta X_{it+k}\theta \tag{3.5}$$

and gap_{it} is estimated as the residuals of equation 3.1,

The empirical results are shown in table 3.12 below. Columns (1) through (4) use the whole sample excluding OPEC countries, whereas columns (5) through (8) restrict it to D10 countries excluding OPEC, and (9) through (12) to OPEC countries only. In each of these cases, the model specification becomes increasingly more general such that columns (1), (5), and (9) are based a DOLS regression of real exchange rate on oil rents per capita. In turn, columns (2), (6), and (10) come from regressions with oil rents per capita and oil price per barrel as explanatory variables, etc.

In almost all specifications and subsamples, we find a negative coefficient that is significantly different from zero at the 5% level on the per capita oil rents variable. A larger coefficient, in absolute value, is observed in the case of OPEC countries. This result corroborates the B-S hypothesis in our sample and suggests that the impact of oil rents per capita on real exchange rate is greater in OPEC countries than the rest of the world. Since these long-run coefficients have an elasticity interpretation, a 10% increase in oil rents per capita in D10 countries leads to a 0.2% appreciation of the currency based on column (8). In the case of OPEC countries, column (12) implies that a 10% increase in per capita oil rents implies approximately 2% appreciation of the currency. However, results for OPEC countries could be spurious as we failed to establish cointegration in the earlier section.

3.6.3 Mean Group and Pooled Mean Group Results

In addition to the DOLS approach, we re-estimate the ECM using Mean Group and Pooled Mean Group estimators. Although not very common within the B-S literature, these estimators exploit the large-N, large-T panel structure effectively. The general model has the following form:

$$lrer_{it} = \sum_{j=1}^{p} \lambda_{ij} lrer_{i,t-j} + \sum_{j=0}^{q} \eta'_{ij} x_{i,t-j} + \mu_i + \varepsilon_{it} \tag{3.6}$$

where x_{it} is a (kx1) vector of explanatory variables for group i, μ_i represent fixed effects, λ_{ij} are scalar coefficients, and η_{ij} is a kx1 vector of coefficients to be estimated. μ_i and ε_{it} denote country fixed effects and the error term, respectively. Assuming $p = q = 1$, this general model can then be re-parametrised into the ECM form as follows:

$$\Delta lrer_{it} = \varphi_i(lrer_{i,t-1} - \beta_i' X_{it}) + \sum_{j=1}^{p-1} \lambda_{ij}^* \Delta lrer_{i,t-j} + \sum_{j=0}^{q-1} \eta_{ij}'^* \Delta X_{i,t-j} + \mu_i + \varepsilon_{it} \tag{3.7}$$

where $\varphi_i = -(1 - \sum_{j=1}^{p} \lambda_{ij})$, $\beta_i = \sum_{j=0}^{q} \eta_{ij} / (1 - \sum_k \lambda_{ik}) = -\sum_{m=j+1}^{p} \lambda_{im}$, with $j = 1, 2, \dots, p-1$, and $\eta_{ij}^* = -\sum_{m=j+1}^{q} \eta_{im}$, with $j = 1, 2, \dots, q-1$.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
loilrents_pc	-0.023***	-0.013**	-0.025***	-0.021***	-0.023***	-0.008	-0.027***	-0.022**	-0.250***	-0.277*	-0.229*	-0.217
	(0.000)	(0.037)	(0.000)	(0.002)	(0.004)	(0.377)	(0.003)	(0.01)	(0.001)	(0.077)	(0.088)	(0.100)
lbrent	—	-0.106***	-0.114***	-0.131***	—	-0.162***	-0.167***	-0.166***	—	0.021	0.136	0.112
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)		(0.931)	(0.543)	(0.608)
lrgdpch	—	—	0.247***	0.125***	—	—	0.316***	0.215***	—	—	-0.752**	-0.598*
			(0.000)	(0.010)			(0.000)	(0.001)			(0.044)	(0.078)
openc	—	—	—	0.004***	—	—	—	0.004***	—	—	—	-0.003
				(0.000)				(0.000)				(0.193)
Speed of adjustment	-0.225***	-0.228***	-0.236***	-0.231***	-0.235***	-0.251***	-0.260***	-0.258***	-0.116***	-0.114***	-0.079	-0.080
Half Lifetime (years)	2.72	2.67	2.57	2.64	2.59	2.39	2.30	2.33	5.62	5.75	8.48	8.36
Subsample	World excl. OPEC				D10 excl. OPEC				OPEC			
N	32				16				10			
Number of observations	819				415				245			

Table 3.12. Dynamic OLS results.

Notes: p-values in parentheses. *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

In most cases, including ours, the parameter of interest is φ_i, which represents the speed of adjustment of the real exchange rate towards the long-run equilibrium. As in all ECM parameterisations, if the long-run relationship exists among the variables, the speed of adjustment is expected to be negative and significant. The parameter vector β_i' contains the long-run coefficients and has an important interpretation in our context.

As noted by Blackburne and Frank (2007) when N and T are large, equation 3.7 can be estimated by a few methods; the spectrum of estimators runs from a fixed effects estimation to Pesaran and Smith's (1995) Mean Group estimator. If the former is implemented, the intercepts are allowed to vary across panel units, but not the slope coefficients. However, FE is consistent only if the slope parameters are homogeneous. If they are not, the MG estimator, which is on the other end of the spectrum from the FE, should be used. MG estimates a separate set of coefficients for each panel unit and then calculates the arithmetic average, which allows all the parameters to vary across countries, including error variances. In our context, this approach has an a priori advantage, since the countries in our dataset have some heterogeneous characteristics. A hybrid approach between these two estimators is Pesaran, Shin, and Smith's (1999) Pooled Mean Group estimator, which pools the long-run coeffeicients accross panels, while averaging intercepts, short-run coefficinets and errror variances. In order to select the appropriate approach between MG and PMG estimator, we use a traditional Hausman test. As usual, if the null hypothesis that the two sets of coefficients are not systematically different is not rejected, PMG is preferred as a more efficient estimator. Otherwise, MG is more appropriate. Table 3.13 shows the results from the models that have been selected by the test.

Dependent variable: *Δlrer*

Specifications	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Hausman (MG vs PMG)	PMG	PMG	DOLS	DOLS	PMG	PMG	DOLS	DOLS	PMG	PMG	DOLS	DOLS
Speed of adjustment	-0.195***	-0.247***	-0.225***	-0.228***	-0.214***	-0.272***	-0.235***	-0.251***	-0.034***	-0.032***	-0.116***	-0.114***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)	(0.007)	(0.009)	(0.006)
Long-run Coefficients												
loilrents_pc	-0.068***	-0.043***	-0.023***	-0.013**	-0.065***	-0.008	-0.023***	-0.008	-1.950***	-2.053***	-0.250***	-0.277*
	(0.000)	(0.000)	(0.000)	(0.037)	(0.000)	(0.200)	(0.004)	(0.377)	(0.000)	(0.001)	(0.001)	(0.077)
lbrent	—	-0.215***	—	-0.106***	—	-0.528***	—	-0.162***	—	-0.062	—	0.021
		(0.000)		(0.000)		(0.000)		(0.000)		(0.953)		(0.931)
Half Lifetime (years)	3.2	2.4	2.72	2.67	2.9	2.2	2.59	2.39	20.3	21.4	5.62	5.75
Subsample	World excl. OPEC				D10 excl. OPEC				OPEC			
N	32				16				10			
N x T	883		819		447		415		265		245	
Log likelihood	817.5	847.0	437.5	439.1	376.4	394.1	431.4	431.7	191.9	197.8	415.8	427.7

Table 3.13. ECM estimation results.

Notes: p-values in parentheses. *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

Focussing on columns (1) through (4), we observe negative and statistically significant speed of adjustment coefficients across all specifications and estimation techniques. These point to the existence of a cointegrating relationship in each specification and provide a robustness check to the cointegration tests discussed and presented in section 3.6.1. Columns (1) and (2) show PMG results, since Hausman tests that MG and PMG results are not systematically different failed to reject the null hypothesis. The estimated long-run coefficients on oil rents per capita and oil price variables are also statistically significant. Using column (2) as an example, we find that a 10% increase in oil rents per capita leads to a 0.4% appreciation of the currency. This provides evidence for the B-S hypothesis in our dataset, but the effect is economically small. Similarly, a 10% increase in the per barrel oil price implies a 2% appreciation of the currency.

Half-life is calculated as $\ln(0.5) / \ln(1 + \varphi)$, which in this specification gives approximately 2.4 years. This implies that we would expect the currency to close half of the gaps between its current level and the long-run equilibrium in about 2.4 years. DOLS results reproduced here are a subset of those presented in table 3.12 and are shown for ease of comparison. Relative to PMG results, DOLS estimates of the long-run parameters are smaller in absolute value. However, the speed of adjustment as well as half-life figures are similar in magnitude. This pattern holds for our smaller D10 subsample, whose results are shown in columns (5) through (8). PMG estimates for D10 countries excluding OPEC members generally show a slightly quicker speed of adjustment: half-lives are calculated as 3.2 and 2.9 years in columns (1) and (5), respectively. This is in agreement with DOLS results, where we observe the same pattern in columns (3) and (7) as well as (4) and (8).

It should be noted that when the oil price is included in the specification for D10 countries, the oil rents per capita variable loses significance – the estimated coefficient becomes closer to zero and the standard error increases dramatically. This implies the existence of a relationship between real exchange rate and the oil price only. Moreover, the long-run coefficients on oil price are fairly large in absolute terms – when PMG is used in (6), the coefficient on the oil price assumes a 5% increase in the value of the currency in response to a 10% increase in the oil price. Furthermore, negative and significant speed of adjustment coefficients point to a cointegrating relationship between real exchange rate and the oil price but not oil rents per capita. This contradicts our findings in columns (5) and (7) as well as columns (1) through (4). Estimating the coefficients in question for the whole sample excluding D10 countries reveals a surprising result: oil rents per capita have a more robust impact on the real exchange rate in countries where oil rents account for less than 10% of GDP in 2008. To see this, table 3.14 shows the results of DOLS regressions for this subsample.

Variables	(13)	(14)	(15)	(16)
loilrents_pc	-0.024*** (0.000)	-0.018*** (0.002)	-0.022*** (0.001)	-0.021*** (0.001)
lbrent	—	-0.057* (0.051)	-0.058* (0.055)	-0.108*** (0.000)
lrgdpch	—	—	0.116* (0.056)	-0.123 (0.101)
openc	—	—	—	0.008 (0.157)
Speed of adjustment	-0.207***	-0.200***	-0.188***	-0.164***
Half Lifetime (years)	2.99	3.11	3.32	3.87
Subsample	World excl. D10			
N	16			
Number of observations	404			

Table 3.14. ECM estimation results for World excluding D10 countries.

Notes: p-values in parentheses. *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

We see here that inclusion of control variables, such as oil price, real GDP per capita, and openness to trade does not have a substantial impact on the coefficient on and statistical significance of oil rents per capita – unlike in the case of D10 countries excluding OPEC.

Lastly, we move on to OPEC output shown in columns (9) through (12) of table 3.13. We interpret these coefficients cautiously, keeping in mind our discussion of OPEC countries in section 3.6.1 on cointegration testing. Firstly, we note that the highly significant speed of adjustment coefficients contradicts our earlier findings using panel cointegration tests: both PMG and DOLS ECM estimations point to the existence of a cointegrating relationship. However, the speed of adjustment coefficients are smaller than in the rest of the sample and imply a half-life of about 6 years when using DOLS and more than 20 years when using PMG. Given that the data is annual and panels on average have 28 observations, these results are meaningless despite a highly significant speed of adjustment estimate in the regression, which is consistent with the results of the Pedroni and Westerlund cointegration tests conducted earlier. Secondly, despite a smaller – in absolute value – speed of adjustment coefficient, we observe a larger – again in absolute value – long-run coefficient on per capita oil rents. PMG results imply elastic long-run coefficients, which suggest that a 10% increase in oil rents per capita leads to approximately 20% appreciation of the currency. Although much smaller, DOLS estimates of long-run coefficients are larger, in absolute value, than DOLS estimates for other subsamples. These coefficients imply a 2.5% appreciation of the currency as a result of a 10% increase in oil rents per capita. Thirdly, and finally, we do not see the pattern we commented on earlier in columns (6) and (8) of the same table. More specifically, adding the per barrel oil price as an additional explanatory variable does not have a notable impact on the coefficient for $loilrents\_pc$, which is again consistent with the Westerlund cointegration test results in section 3.6.1. As a final comment, we note the number of panels is small for both OPEC and D10 excluding OPEC subsamples. Given the asymptotic properties of PMG estimator and the requirement for both N and $T \rightarrow \infty$, we acknowledge the unreliability of these estimates and put more weight on DOLS results. We include PMG results as a robustness check as opposed to a conclusive estimate.

In the output table 3.13 above, we have opted for a smaller model specifications consisting of per capita oil rents and oil price. This is mainly due to the behaviour of the GDP variable and the correlation between the variables. To see this, we turn to table 3.15 below, which summarises the MG and PMG models estimated using our largest subsample.

Although we observe a negative and statistically significant speed of adjustment coefficient across all specifications, long-run coefficients behave unexpectedly with larger model specifications. Specifically, comparing columns (2) and (3), we observe that the Hausman test null hypothesis is rejected when GDP per capita is included as an explanatory variable. This suggests that the long-run coefficient on GDP per capita is heterogeneous across countries unlike oil rents per capita and oil price. Estimating a simple specification with GDP per capita as the only explanatory variable confirmed this observation. Furthermore, the link between oil rents per capita, oil price, and GDP per capita leads to inverted signs on the coefficients of $loilrents\_pc$ and $lrgdpch$. Even though none of these long-run coefficients are statistically significant, this finding is surprising. Based on Balassa's (1964) original work and our findings elsewhere in this chapter, we would expect a negative and significant coefficient. Given that the speed of adjustment coefficient is negative and significantly different from zero, this behaviour could be attributed to the correlation between the three explanatory variables in specification (3). Calculating a simple correlation matrix shows that the correlation between $loilrents\_pc$ and $lrgdpch$ is approximately 0.3, which appears to be sufficiently high to cause misleading results. Due to these observations, we have opted for smaller model specifications.

Dependent variable: Δlrer

Specifications	(1)	(2)	(3)	(4)
Hausman (MG vs PMG)	PMG	PMG	MG	MG
Speed of adjustment	-0.195*** (0.000)	-0.247*** (0.000)	-0.418*** (0.000)	-0.484*** (0.000)
Long-run Coefficients				
loilrents_pc	-0.068*** (0.000)	-0.043*** (0.000)	0.088 (0.415)	0.506 (0.318)
lbrent	—	-0.215*** (0.000)	-0.207 (0.192)	-1.048 (0.223)
lrgdpch	—	—	0.047 (0.920)	0.093 (0.834)
openc	—	—	—	0.014** (0.019)
Half Lifetime (years)	3.2	2.4	1.3	1.0
Subsample	World excl. OPEC			
N	32			
N x T	883			
Log likelihood	817.5118	846.9928	1114.142	—

Table 3.15. Comparison of the MG and PMG estimation results.

Notes: p-values in parentheses. *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

3.6.4 Further investigation of OPEC countries.

The real exchange rates in OPEC countries appear to behave differently from the rest of the sample – they didn't exhibit any cointegrating relationship with oil rents per capita. In order to explain the lack of a long-run relationship, a country-by-country cointegration tests have been performed. We opted for the traditional Engle and Granger (1987) cointegration test, which applies an ADF test to the residuals from a static regression of the real exchange rate on the oil rents per capita. The critical values differ from the standard ADF test and are provided by Mackinnon (2010). The optimal lag length for the test was chosen according to Ng and Perron (1995), but it should be noted that the outcomes are robust to the lag length selection. The results are presented in table 3.16. As shown in the table, no cointegrating relationship was found for any of OPEC countries. Further investigations revealed that real exchange rate and the oil price are not cointegrated either (the only exception being Ecuador). If all three variables (oil price, oil rents, and the real exchange rate) are used in the same long-run equation, there is no cointegration found for any OPEC country.

Country	Optimal lag length	Test statistic	1% cr.value	5% cr.value	10% cr.value	Number of observations
Angola	0	-1.911	-4.415	-3.615	-3.234	23
Algeria	0	-1.85	-4.157	-3.479	-3.142	44
Ecuador	1	-2.19	-4.966	-3.893	-3.417	11
Iraq	0	-1.508	-4.192	-3.497	-3.155	39
Libya	0	-1.738	-4.415	-3.615	-3.234	23
Nigeria	0	-2.338	-4.228	-3.516	-3.168	35
Qatar	4	-2.779	-4.415	-3.615	-3.234	19
Saudi Arabia	2	-1.209	-4.415	-3.615	-3.234	21
UAE	5	-2.169	-4.415	-3.615	-3.234	18
Venezuela	0	-0.518	-4.5	-3.659	-3.263	20

Table 3.16. Individual countries (OPEC) cointegration tests

One potential explanation for the difference in the behaviour of OPEC countries from the rest of the oil exporters in the sample could be linked to countries' currency regimes. Appreciation of the real exchange rate can happen if either the nominal exchange appreciates or the price level increases. The latter change, when caused by an increase in productivity of tradables, is the mechanism behind the B-S hypothesis.

Devereux (2014) notes that the nominal exchange rate fluctuations introduce noise into the estimation of the B-S effect as they tend to change more rapidly compared to the price level. In practice, most OPEC countries don't have free floating currency regimes. Saudi Arabia, Qatar and the United Arab Emirates all peg their currencies to the US dollar and have done so throughout the estimation period. Venezuela pegged their currency to the US dollar in 2003. Ecuador has adopted the US dollar as an official currency in 2000 after almost two decades of a crawling peg. Iraq's currency regime is a managed float, however the official rate has been pegged to the US dollar at various times. Libya's currency is pegged to a composite exchange rate anchor. If a currency is pegged it should be easier to separate changes in the real exchange rate attributable to the adjustments in the price level compared to the adjustments of the nominal exchange rate as the numerator of the real exchange rate stays fixed. It is possible that the long-run relationship found in D10-OPEC countries is driven by the nominal exchange rate rather than the price level adjustments, so the absence of the B-S mechanism is just more prominent in OPEC countries. This is exacerbated by the fact that the nominal exchange rate in some oil-exporting countries (for example, Russia) will be highly responsive to the oil price fluctuations. This phenomenon will be most evident in countries where oil represents large fraction of the export share. For example, 83% of Sudan's 2009 exports (Sudan is one of the countries with cointegrating relationship between oil rents and the real exchange rate) consisted of crude oil and another 2% of refined petroleum oil according to "The Atlas of Economic Complexity". So it is possible that the relationship detected earlier is not driven by the B-S mechanism, but is driven by the nominal appreciation of the currency in response to the appreciation of its exports, demand for which is inelastic.

In order to test the proposition above Westerlund cointegration tests (described in section 3.6.1) have been conducted using log of nominal exchange rate (lxrat) and log of PPP price level (lppp) as dependent variables and oil rents per capita as an independent variable. The results, presented in table 3.17, negate the proposition – there is strong evidence that in non-OPEC countries unlike OPEC countries, both, nominal exchange rate and price level are cointegrated with the oil sector productivity, while neither of the two are cointegrated with the per capita oil rents in OPEC countries.

Dependent variable	Test statistic		(1)	(2)	(3)
lxrat	Pt	z-value	-1.345*	-16.5***	-30.8***
		p-value	0.089	0.000	0.000
lxrat	Pa	z-value	1.2	-8.2***	-10.4***
		p-value	0.885	0.000	0.000
lppp	Pt	z-value	-0.5	-21.0***	-42.8***
		p-value	0.310	0.000	0.000
lppp	Pa	z-value	1.315	-12.7***	-18.9***
		p-value	0.906	0.000	0.000
	Subsample		OPEC	World-OPEC	D10-OPEC
	N		10	32	16
	Lags		1	1	1

Table 3.17. Westerlund (2007) cointegration tests.

Note: *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

Another potential explanation is that some assumptions of the B-S hypothesis are failing in the context of OPEC countries. It can be seen in figure 3.14 that most OPEC countries appear to have lower price level compared to other countries with similar per capita incomes, which suggests that the price level is not adjusting normally as productivity in tradables increases. In particular, note Qatar, Saudi Arabia, UAE and Libya – all four countries have lower-than-expected price levels, and all four countries are characterised by migration policies which admit large numbers of low-wage temporary workers.[12] These migrants act to depress both wages and productivity in the non-tradable sector relative to a country like Norway. Effectively, these countries undermine assumption of the B-S hypothesis that the workers can freely move between tradable and non-tradable sector, so wages in the non-tradable sector simply do not adjust or don't fully adjust to wages in tradable sector and the overall price level stays low.

[12] For Libya, we are of course referring to the pre-civil war era contained in our data set. "Migrant workers make up the majority of the population in Bahrain, Oman, Qatar and the United Arab Emirates (and more than 80 per cent of the population in Qatar and the United Arab Emirates); while in construction and domestic work in Gulf States, migrant workers make up over 95 per cent of the work force." (Labour Migration (Arab States), International Labor Organization, Accessed 08/04/2017, http://www.ilo.org/beirut/areasofwork/labour-migration/lang--en/index.htm)).

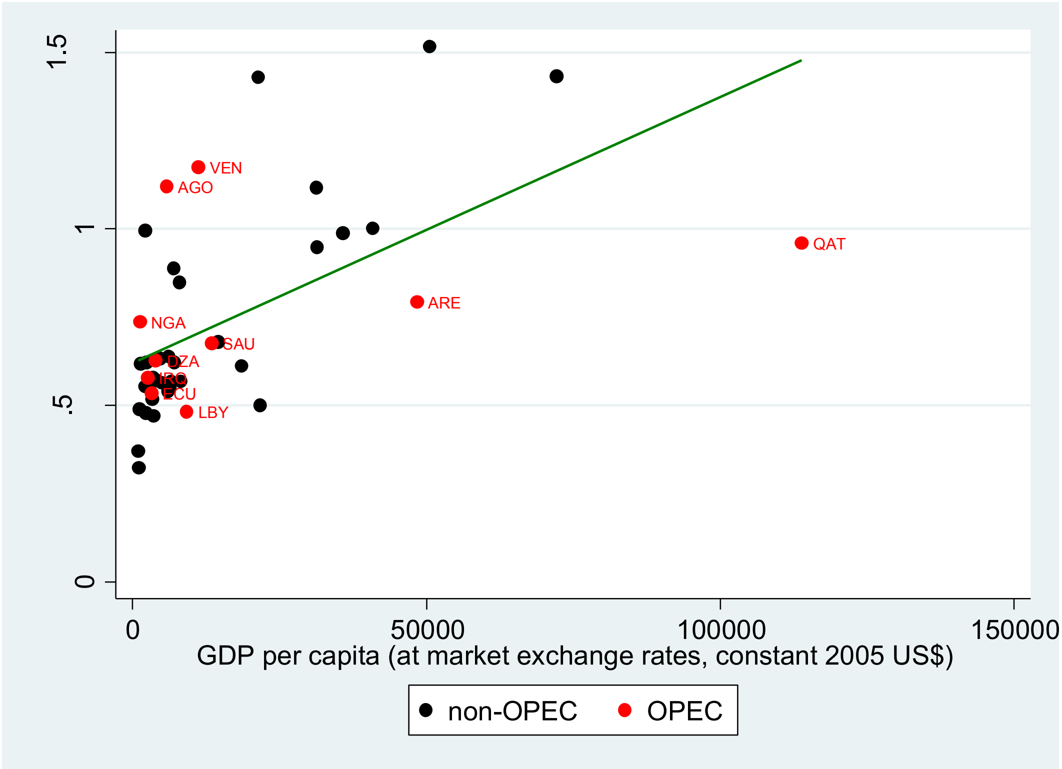

Figure 3.14. Log of per capita income and PPP ratio in OPEC vs. non-OPEC countries in 2009

The behaviour described above combined with low power properties of individual cointegration tests means that it is in principle hard to reject the null of no cointegration for an individual country even when behaviour does follow the B-S effect and when countries exhibit different behaviour we reject OPEC as a whole panel.

3.7 Conclusion.

In this chapter an attempt has been made to identify whether changes in oil sector productivity can explain real exchange rate volatility in oil-exporting countries, which in turn could explain whether the countries are affected by the Dutch disease or not. The analysis was based on the unique dataset that allowed us to calculate profits made in the oil sectors in several countries over time and separate the effect of an increase in productivity of the oil sector as opposed to the traditional use of GDP as a proxy for productivity of the tradables in the context of the B-S hypothesis. If the effect is identified, it carries importance for policy-making, especially in developed countries as noted by Chen and Rogoff (2002).

We found that the B-S mechanism holds in some oil-exporting countries, but doesn't in the others. The relationship was the strongest in oil-exporting countries where oil sector productivity constitutes small fraction of the country's GDP (World-D10 countries) and was highly significant for most specifications for the D10-OPEC countries. For all countries that are not members of the OPEC, the relationship has been found to be significant and the sign is always as expected, however, the magnitude of the effect is tiny. The largest observed effect is a 0.7% increase in the real exchange rate in response to a 10% in the oil sector productivity. At the same time, the effect of the oil price changes on the real exchange rates is up to 8 times larger than the effect caused by changes in productivity.

In OPEC countries no evidence was found for the B-S mechanism – an increase in oil sector productivity doesn't appear to have any effect on either the nominal exchange rate (it has to be noted that most countries have a nominal exchange rate peg) nor on the price level. Moreover, unlike Korhonnen and Juurikkala (2009), the real oil price proved to have no effect on the real exchange rates of the OPEC countries either. In section 3.6.4 we have discussed a potential explanation for this unexpected result – at least one of the assumptions of the B-S hypothesis, free movement of labour between tradable and non-tradable sector, appears to be failing in some OPEC countries.

Our overall conclusion is that the B-S hypothesis is likely to hold in most oil-exporting countries, however, the magnitude of the effect appears to be very small. This suggests that other potential explanations of the resource curse should be considered. Our results are consistent with Van der Ploeg (2011), who found that even though the B-S mechanism is responsible for the resource curse to a certain extent, the main contributors are corruption, low quality of institutions, and underdeveloped financial systems that fail to mitigate the high volatility of commodity prices. Also, if the deviations from PPP are modelled, it wouldn't help much to use oil productivity to explain the deviations. Future work might revisit this question using data with more observations and/or higher frequency.

Chapter 4. How Costly is Conservation? The International Energy-GDP Relationship Re-examined

4.1 Introduction

Does economic growth lead to increased energy consumption? Does increased energy consumption lead to economic growth? There are four possibilities: causality is bi-directional (known as the feedback hypothesis), i.e. energy consumption and GDP drive each other; it is unidirectional with energy consumption driving economic growth (the growth hypothesis); or vice versa (the conservation hypothesis); and the last one assumes no relationship between the two (neutrality hypothesis).

	GDP ⇒ Consumption	*GDP ⇏ Consumption*
Consumption ⇒ GDP	*Feedback hypothesis*	*Growth hypothesis*
Consumption ⇏ GDP	*Conservation hypothesis*	*Neutrality hypothesis*

Table 4.1. Four possible causal relationships for energy consumption and GDP

These four hypotheses have been widely tested in the literature in order to identify the nature of the relationship, but there has been little consensus. The answer to the question, however, has important policy implications. In setting climate policy to deal with carbon emissions, for example, it is important to know how much of an effect reduced energy consumption might have on economic growth, if indeed there is any effect at all. Or consider energy shocks: what is the effect on the economy of a sudden change in energy consumption, as might happen after a natural disaster (e.g. Japan's 2011 Fukushima earthquake and tsunami, leading to a national shutdown of nuclear power stations), or a political crisis (e.g. any of various Middle Eastern wars and civil wars over the past 40 years)?

Earlier studies in the field approached this issue using individual country-level time series regressions, which produced conflicting results. Those early studies relied on the datasets that spanned short time periods and were likely to suffer from small sample problems and produce deceptive results, so in more recent years panel econometric methods gained popularity as they could potentially enhance the individual time series approaches by pooling data across countries to improve efficiency of estimations, which in turn would improve power properties of many tests involved in the analysis. However, little emphasis is given to how we should interpret the results of these panel estimations and whether these more efficient estimations are actually meaningful.

In this chapter I re-investigate the relationship between energy consumption, economic growth and energy prices using a recent dataset (1978-2013) of 28 OECD countries and up-to-date econometric methods (details below) suitable for this analysis. I also focus on how we should think about panel estimations in the context of energy-GDP cointegration as it appears that most studies adopt panel estimations without questioning the underlying assumptions that are made when moving from country level analysis to panel estimations, which might or might not be true. I argue that first, cross sectional dependence is very likely to be present in macro panels, so it has to be accounted for properly, and I also show that panels are likely to be heterogeneous in terms of the relationship between energy consumption and GDP, which should also be taken into account.

The contributions of the present work to the literature are several: 1. I use a longer time series that incorporates the most recent financial crisis (1978-2013). 2. I investigate the relationship in all 28 OECD countries (typically a smaller subset of the OECD countries is used for the analysis). 3. I use a new framework for unit root testing developed by Bai and Ng (2004, 2010) that is robust to cross-sectional dependence. 4. I employ Westerlund's (2007) cointegration test that accounts for cross-sectional dependence by bootstrapping standard errors. 5. I estimate the relationship and the direction of causality using Mean Group (MG) (Pesaran and Smith, 1995), Pooled Mean Group (PMG) (Pesaran et al., 1999) and Common Correlated Effects Mean Group (CCEMG) (Pesaran, 2006) estimations, compare the results, and make suggestions for the future panel-based research of the energy-growth nexus.

Based on the methods mentioned above, I find evidence mainly for the *conservation* hypothesis - economic growth drives energy consumption in the long run, but not the other way around, which suggests that modest energy conservation policies should not affect economic growth adversely.

The remainder of this chapter is organised as follows: section 4.2 provides a brief overview of the previous studies, section 4.3 describes the data, econometric methods and explains the results of the estimations, and section 4.4 concludes the analysis.

4.2 An overview of the literature and a summary of recent developments

The pioneer study that investigated the relationship between energy consumption and economic growth was Kraft and Kraft (1978), which also found support for the conservation hypothesis; the authors concluded that GNP drove energy consumption in the post-war period in the US, but they did not find evidence that energy consumption drove GNP. Subsequently, many studies have re-examined the relationship using different methods or different time periods and found conflicting results. For example, Akarca and Long (1980) and Eden and Hwang (1984) concluded in favour of the neutrality hypothesis that there is no causal relationship in either direction; Stern (2000) reported that in most specifications he considered for US data, he found support for the growth hypothesis of unidirectional causality from energy consumption to GDP, while Lee (2006) found support for the feedback hypothesis of mutual casualty between energy consumption and the economy at large. So not only have all four possibilities found empirical support in the literature, but all four possibilities have found support when considering exclusively US data, which is perhaps surprising.

Given that all four possibilities have found support in US time series data, the reader may not be surprised to discover that the findings from time series data in an international context are also somewhat varied. Masih and Masih (1997) found support for the bidirectional feedback hypothesis in Taiwan and South Korea. Asafu-Adjaye (2000) established long-run unidirectional causality from energy and the consumer price index to economic growth (the growth hypothesis) in India and Indonesia, but found support for the feedback hypothesis for Thailand and the Philippines. Soytas and Sari (2003) concluded in favour of the neutrality hypothesis for the UK, US and Canada.

More recent studies have used applied panel methods to analyse the GDP-energy relationship. Al-Iriani (2006) considered the energy and economic growth relationship for the six Gulf Cooperation Council countries[13] from 1971–2002 using panel methods to test for cointegration and the direction of causality and found that the causality runs from GDP to energy consumption, but not conversely (i.e. the conservation hypothesis). Huang et al. (2008) conducted a comprehensive study of 82 countries in the period from 1972 to 2002 – they split the countries into low, middle and high income countries and concluded that GDP was positively related to energy consumption in middle income countries, while high income countries exhibit a negative relationship. No relationship was found for the low income countries. Apergis and Payne published several studies (including 2009, 2010a, 2010b, 2012) that analysed different groups of countries (11 ex-Soviet countries of the Commonwealth of Independent States in the first two studies, nine South American countries in the third, and 80 countries in the fourth) using similar econometric methods – first the Pedroni panel cointegration tests are used to establish the presence of a long-run relationship between the variables, and then a panel error correction model (ECM) is used to determine the direction of causality. These methods have been widely used in the literature (for example, Ozturk et al. (2010), or Streimikiene and Kasperowicz (2016)). I will argue in Section 3 below that both Pedroni, which only deals with simple cross-sectional dependence, and panel ECM, which assumes homogeneity of parameters across panels, may not be appropriate methods in the context of energy – GDP relationship analysis.

For a broader view of this literature, both Ozturk (2010) and Coers and Sanders (2013) provide excellent overviews. Coers and Sanders (2013) also explain the historical development of the methods. They split all the existing studies into five generations, with traditional vector autoregressive model (VAR) analysis being the first generation, bi-variate ECM being second, multivariate cointegration analysis (Johansen's method) being third, then panel cointegration analysis with few panel units are fourth, and the fifth and final generation are studies that use a wide range of countries in the context of panel estimation.

There are a couple of additional studies worth highlighting. Belke et al. (2011) analyse the relationship between energy consumption, income and energy prices for a panel of countries by decomposing each variables into common and idiosyncratic factors by means of principal component analysis (PCA). The interpretation of the common

[13] Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates.

factors in their study corresponds to the variation in the data, which is common across countries, while the idiosyncratic components correspond to country-specific factors. So for example the common component for GDP would correspond to something like the global business cycle, while the idiosyncratic factor would represent a country-level deviation from global trends. Belke et al. conclude that the common factors are cointegrated, while idiosyncratic factors are stationary. This means that the idiosyncratic factors cannot participate in the long-run cointegrating relationship. The findings support the feedback hypothesis of mutual causality between energy and GDP. They also suggest that country-level energy policies cannot have significant effects, as the relationship is driven by the common international trends. One shortcoming of this study concerns the estimators which were used to evaluate the long-run relationship and the direction of causality – dynamic ordinary least squares (DOLS) (Mark and Sul, 2003) and the Arellano Bond (1991) estimator. Neither estimator allows for heterogeneity across panels, whereas it is hard to imagine the same parameters characterising Norway and Greece.

Another study that takes advantage of newly developed panel methods is Damette and Seghir (2013). They use the Westerlund (2007) cointegration test to test for the presence of a long-run relationship between energy consumption and GDP, and use the pooled mean group estimator (PMG) of Pesaran et al. (1999) to determine the direction of causality. Damette and Seghir (2013) analysed 12 oil exporting countries and established that in the short run, energy affects income, but in the long run it's economic growth that drives energy consumption. They do not, however, test whether the PMG estimator is likely to be consistent in the context of their analysis.

In contrast to these two studies, I will show that when a longer period is considered, not only common factors, but also idiosyncratic factors are non-stationary, so country-level policies could be influential. I will also show that assuming homogeneity of parameters across panels is restrictive, and not supported by the data.

4.3 Data, methodology an empirical results

4.3.1 Data

GDP per capita (*gdp_pc*) is measured in 2005 PPP USD, total final energy consumption per capita (*tfc_pc*) is measured in tonnes of oil equivalent, and energy prices are measured with an index of the real end-use energy price for industry and households (*ritotal*), which is normalised to a 100 in 2010 for each country. All data come from the International Energy Agency's (IEA) – *gdp_pc* and *tfc_pc* are taken from the IEA's World Energy Balances, while the price index data comes from the IEA's quarterly edition on Energy prices and taxes. All variables are converted to natural logs.

The data consist of a blanced panel of 28 OECD countries covering the years from 1978 to 2013 inclusive. The countries consist of : Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, South Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, the UK and the USA. Other OECD countries - Chile, Estonia, Latvia, Iceland, Israel, Slovenia and Turkey - have been excluded as data are partially absent and their inclusion would unbalance the panel. I also consider separately the results for the G7 countries: Canada, France, Germany, Italy, Japan, the UK and the USA.

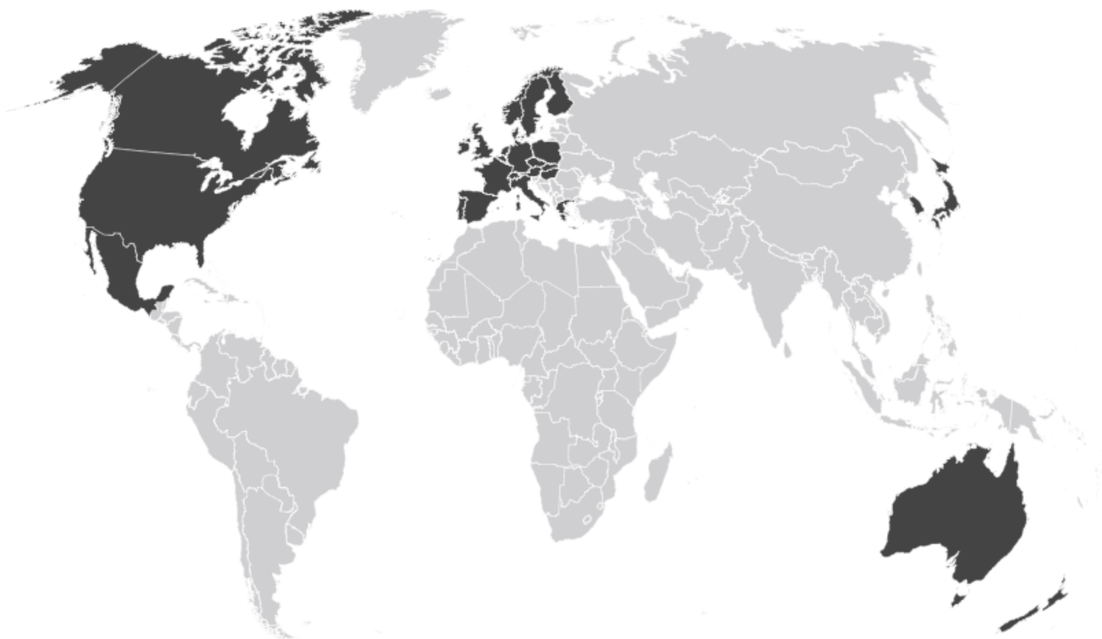

Figure 4.1. OECD sample of countries (see text for full list).

All variables are illustrated in figures A.1-A.3 of the Appendix.

4.3.2 Common and idiosyncratic factors

Bai and Ng (2004) developed a general framework (referred to as PANIC – Panel Analysis of Non-stationarity in Idiosyncratic and Common components) to test for unit roots in large-N large-T panel data with potential cross-sectional dependence, which is likely to be present in macro panels. When not dealt with properly, cross-sectional dependence results in deceptive inference both when the unit root tests are performed for panels as a whole and also when the individual series are tested and then individual statistics are pooled (Reese and Westerlund, 2015). Bai and Ng's (2004) approach hinges on the idea that the series under examination can be split into common trends, i.e. co-movements of the data in all panel units over time and idiosyncratic factors, movements in the data that are specific to individual panel units. It is based on the application of principal component analysis (PCA) to the original data, which separates common and idiosyncratic trends in the data and tests them separately. After the common factors have been separated from the variables, the remaining data should only exhibit low levels of cross-sectional dependence. If both common and idiosyncratic components are found to be I(0), the original series is considered to be I(0), but if either or both of the common and idiosyncratic factors are found to be I(1), the original series necessarily has to be I(1). Another benefit of this method is that it doesn't require pre-testing data for the order of integration and can deal with all possibilities i.e. both common factors and idiosyncratic factors are I(1), both are I(0), either is I(1) while the second one is I(0). The method works by differencing the original series before applying the PCA and then restoring the integration properties of the series by recumulating the differenced factors. In addition to solving problems arising from cross-sectional dependence, the decomposition has an economic meaning - it allows researchers to analyse international trends separately from country-specific movements and variation.

Here, I opt for correlation-based PCA, where variances of all variables are normalized before the decomposition. The effect of this choice is to put similar weights on the variation in variables in richer and poorer countries. But the choice is not likely to make much difference in any case as all the variables are already natural logs. The results of the PCA decomposition, presented in table 4.2, suggest that all three variables are well explained by the first two principal components and they account for at least 78% of variation in the panel data.

The Johansen (1988) cointegration test is used to investigate whether the common components are cointegrated. Due to small number of observations the lag length is chosen by examining properties of unrestricted VARs[14] with one and two lags and a one lag specification is chosen based on satisfactory post estimation analysis – the residuals appear to be normal and, despite small evidence of autocorrelation at lag one, it disappears at all subsequent lags. The vector error correction model (VECM) is then estimated, based on the underlying VAR with one lag, and the maximum eigenvalue and the trace statistic are used to determine the number of cointegrating relations.

Variable	**Component**	**Eigenvalue**	**Proportion of variation explained**	**Cumulative proportion of variation**
	Comp1	26.0	0.93	0.93
gdp_pc	Comp2	1.3	0.05	0.97
	Comp3	0.3	0.01	0.98
	Comp1	15.9	0.57	0.57
tfc_pc	Comp2	5.8	0.21	0.78
	Comp3	2.0	0.07	0.85
	Comp1	18.9	0.67	0.67
ritotal	Comp2	5.8	0.21	0.88
	Comp3	1.3	0.05	0.93

Table 4.2. Principal component analysis of GDP per capita, energy consumption per capita and the real price index.

All six principal components are tested for stationarity using Elliot, Rothenberg and Stock's (1996) (ERS) DF-GLS test, which is a more powerful modification of the ADF test. The lag length is determined based on the MAIC, the modified Akaike information criterion (Ng and Perron, 2000). It can be seen in table 4.3 that all the common components are non-stationary around a linear trend, but their first differences are stationary, so the original common components are I(1).

[14] Small sample degree of freedom adjustments are made

Variable	№ of obs.	Lags	Trend	Test statistic	Rejection	Stationary or not
pca_gdp_pc_1	34	1	Yes	-2.056	Do not reject at 10% s.l.	Non-stationary
pca_gdp_pc_2	34	1	Yes	-1.439	Do not reject at 10% s.l.	Non-stationary
pca_tfc_pc_1	33	2	Yes	-1.281	Do not reject at 10% s.l.	Non-stationary
pca_tfc_pc_2	34	1	Yes	-2.284	Do not reject at 10% s.l.	Non-stationary
pca_ritotal_1	34	1	Yes	-1.584	Do not reject at 10% s.l.	Non-stationary
pca_ritotal_2	34	1	Yes	-1.238	Do not reject at 10% s.l.	Non-stationary
d. pca_gdp_pc_1	31	2	No	-2.039	Reject at 5% s.l.	Stationary
d. pca_gdp_pc_2	32	1	No	-1.714	Reject at 10% s.l.	Stationary
d.pca_tfc_pc_1	32	2	No	-1.671	Reject at 10% s.l.	Stationary
d.pca_tfc_pc_2	32	2	No	-3.765	Reject at 1% s.l.	Stationary
d.pca_ritotal_1	32	2	No	-2.768	Reject at 1% s.l.	Stationary
d.pca_ritotal_2	33	1	No	-2.921	Reject at 1% s.l.	Stationary

Table 4.3. First two principal components of GDP per capita, energy consumption per capita and the real price index and their first differences unit root test results

The maximum eigenvalue statistic rejects the null hypothesis of 'no cointegration' in favour of one cointegrating relation. The trace statistic, on the other hand, suggests that there are four cointegrating relations among the common factors. The results are presented in table 4.4. The disagreement between these two results is not investigated further as the main point at this stage is to determine whether common factors are cointegrated or not, and it appears that they are.

Maximum rank	Trace statistic	5% critical value	Maximum-eigenvalue statistic	5% critical value
0	145.10	94.15	55.40**	39.37
1	89.69	68.52	31.56	33.46
2	58.13	47.21	26.58	27.07
3	31.55**	29.68	16.91	20.97
4	14.63	15.41	12.50	14.07
5	2.14	3.76	2.14	3.76

Table 4.4. Cointegration test of the common factors

Having established a cointegrating relationship in the levels of the common factors of all three variables, we move to testing idiosyncratic errors for stationarity. If the idiosyncratic errors are stationary, this would suggest that the long-run relationship between the variables exists and is driven by the common movement across panels. The idiosyncratic components are constructed as a residual from a regression of the differenced variables on their differenced first two common factors. The residuals are then cumulated up to reconstruct the initial order of integration of the data in levels. I opt for the Pa, Pb and PMBS tests developed by Bai and Ng (2004, 2011) specifically for PANIC residuals and are shown to have good finite sample properties. All tests have a null of a unit root in all panels and the critical values are derived using Monte Carlo simulations (Bai and Ng, 2011). The results, shown in table 4.5, suggest that the idiosyncratic components of GDP per capita, energy consumption per capita and energy prices are non-stationary as for none of the test statistics the null can be rejected at 5% significance level. This result were supported by standard panel unit root tests – Levin, Lin and Chu (2002) and Harris and Tzavlis (1999).

Test	**Idiosyncratic component of GDP per capita**		**Idiosyncratic component of energy consumption per capita**		**Idiosyncratic component of energy price index**	
	Test statistic	**p-value**	**Test statistic**	**p-value**	**Test statistic**	**p-value**
Pa	-1.254*	0.10	0.902	0.82	-1.269*	0.10
Pb	-0.856	0.20	0.982	0.84	-1.146	0.13
PMSB	-1.545*	0.06	0.908	0.82	-0.528	0.30

Table 4.5. Unit root tests of the idiosyncratic components

Notes: *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

The results conflict with Belke et al. (2011), who found all idiosyncratic components to be I(0) and concluded that the cointegrating relationship between GDP, energy consumption and energy prices must be driven by the relationship in the common trend of OECD counties, which resulted in a position that the efficacy of national policies is overshadowed by international factors.

The next step is testing the original variables for cointegration. I established earlier that common factors are I(1) and cointegrated; however, idiosyncratic factors were also found to be non-stationary, so it is necessary to test whether the long-run relationship exists when both common and idiosyncratic trends are taken into account.

4.3.3 Testing for panel cointegration

Most studies use the Pedroni (1999, 2004) residual based cointegration test to check for panel cointegration in the context of the relationship between energy and economic growth (e.g. Al-Iriani (2006), Apergis and Payne (2009, 2010a, 2010b, 2012) Streimikiene and Kasperowicz (2016), etc.). But the Pedroni test assumes errors are uncorrelated across panels, which is very likely to be violated here, and can account only for simple cross-sectional dependence by including time dummies into the estimated equations. I opt instead for Westerlund's (2007) cointegration test with bootstrapped standard errors which make the test robust to cross-sectional dependence. The test estimates the ECM-type equation and checks whether the panels are error-correcting towards the long-run equilibrium. One caveat with this test is that the results based on the ECM will vary depending on which variable is chosen as a dependent variable. Up to this point, all the tests considered in this analysis didn't require an imposition of the direction of causality, which itself is one of the questions under investigation. To allow for the various causal relationships, it is necessary to perform the test three times, using each possible variable as the dependent variable in turn (equations 4.1-4.3). For example, if equation 4.1 is estimated with GDP per capita as the dependent variable, a significant speed of adjustment coefficient φ_1 would imply that GDP is adjusting towards a long-run equilibrium among the three variables. If, on the other hand, φ_1 is not found to be statistically significant, this does *not* imply that there is no long-run relationship; it merely implies that GDP is not the variable which is adjusting towards the long-run relationship. The long-run equilibrium relationship may or may not exist, but if it does exist, it is one or both of the other variables (in this case, energy consumption or prices) which are adjusting themselves to equilibrium.

The equations are as follows:

$\Delta gdp_{pc_{it}} = \varphi_{1i}(gdp_{pc_{it-1}} - \beta_{11i}tfc_{pc_{it-1}} - \beta_{12i}ritotal_{it-1}) + \alpha_{1i} +$
$\sum_{k=0}^{p} \Delta\, tfc_{pc_{it-k}}\theta_{11i} + \sum_{k=0}^{p} \Delta\, ritotal_{it-k}\theta_{12i} + \sum_{k=1}^{p} \Delta\, gdp_{pc_{it-p}}\theta_{13i} + v_{1it}$
(4.1)

$\Delta tfc\_pc_{it} = \varphi_{2i}(tfc\_pc_{it-1} - \beta_{21i}gdp\_pc_{it-1} - \beta_{22i}ritotal_{it-1}) + \alpha_{2i} +$
$\sum_{k=0}^{p} \Delta\, gdp\_pc_{it-k}\theta_{21i} + \sum_{k=0}^{p} \Delta\, ritotal_{it-k}\theta_{22i} + \sum_{k=1}^{p} \Delta\, tfc\_pc_{it-p}\theta_{23i} + v_{2it}$
(4.2)

$\Delta ritotal_{it} = \varphi_{3i}(ritotal_{it-1} - \beta_{31i}gdp\_pc_{it-1} - \beta_{32i}tfc\_pc_{it-1}) + \alpha_{3i} +$
$\sum_{k=0}^{p} \Delta\, gdp\_pc_{it-k}\theta_{21i} + \sum_{k=0}^{p} \Delta\, tfc\_pc_{it-k}\theta_{22i} + \sum_{k=1}^{p} \Delta\, ritotal_{it-p}\theta_{23i} + v_{3it}$
(4.3)

where i subscripts denote countries, t subscripts denote years, φ_i represents the speed of adjustment coefficients in country i, α_is denotes the country fixed effects, θ_{3i} are $(p \times 1)$ vectors of coefficients on the lagged first-differenced dependent variable and θ_{1i} and θ_{2i} are $\left((p+1) \times 1\right)$ vectors of coefficients on the lagged first-differenced regressors. The error terms are denoted as v_{it}.

Dependent	**Lags and**	**Test statistic**	**Gt**	**Ga**	**Pt**	**Pa**
gdp_pc	3	z-value	0.025	3.559	-0.398***	0.915**
		robust p-value	0.37	0.24	0	0.02
tfc_pc	2	z-value	-2.093**	0.974	-2.002*	-0.224
		robust p-value	0.047	0.453	0.053	0.247
ritotal	2	z-value	-3.428**	1.042	-2.23**	-0.735
		robust p-value	0.02	0.433	0.033	0.253

Table 4.6. Westerlund's (2007) cointegration test results

Notes: *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

Westerlund's (2007) cointegration test consists of four test statistics - Gt, Ga, Pt and Pa. Gt and Ga are based on a weighted average of φ_i obtained from different panel estimations, while Pt and Pa pool the data across panels to obtain an estimate of φ for the whole panel. All test statistics have a null of no cointegration with the alternatives of "at least one panel contains cointegration" (if Gt and Ga are used) or the alternative of "all panels contain cointegration" (Pa and Pt statistics). The optimal leg length is selected according to the Akaike information criterion (AIC). To account for potential cross-sectional dependences bootstrapped critical values are used and robust p-values are reported. The results of the test, presented in table 4.6, are ambiguous. For each dependent variable, two of the test statistics reject the null of no cointegration at the 10% level, while the other two do not reject the null. The ambiguity could be attributed to a small number of observations per panel – as an ECM-based test, the Westerlund test suffers when the number of time periods is small, especially if the number of lags and leads is non-zero. However, as half of the test statistics indicate cointegration, I will assume that the long-run relationship exists and will further explore this issue in the next section.

4.3.4 Long-run relationship (DOLS).

The results in the previous section indicate that GDP per capita, energy consumption per capita and the real energy price index appear to share a common stochastic trend, so the next step is to estimate the relationship – evaluate magnitudes of the coefficients, test which variables are adjusting towards the long-run relationship and how quickly they do so. I will start with a panel version of Stock and Watson's (1993) dynamic ordinary least squares estimator (DOLS), which is a consistent and asymptotically efficient estimator of the cointegrating vector and, as noted by Kao and Chiang (2001) and Mark and Sul (2003), outperforms panel OLS and fully modified least squares estimator (FMOLS) by having a smaller bias and smaller finite sample size distortions. The estimator requires inclusion of leads and lags of the differenced regressors into the estimated equation in order to make the error term orthogonal to stochastic shocks in the regressors. To obtain asymptotically valid standard errors, a heteoskedasticity- and autocorrelation-robust (Newey-West) estimator will be used.

The DOLS equations have the following form:

$$gdp\_pc_{it} = \alpha_{1i} + tfc\_pc_{it}\,\beta_{11} + ritotal_{it}\beta_{12} + \sum_{k=-a}^{b} \Delta\, tfc\_pc_{it+k}\theta_{11} + \sum_{k=-a}^{b} \Delta\, ritotal_{it+k}\theta_{12} + \epsilon_{it} \quad (4.4)$$

$$tfc\_pc_{it} = \alpha_{2i} + gdp\_pc_{it}\,\beta_{21} + ritotal_{it}\beta_{22} + \sum_{k=-a}^{b}\Delta\, gdp\_pc_{it+k}\theta_{21} + \sum_{k=-a}^{b}\Delta\, ritotal_{it+k}\theta_{22} + \varepsilon_{it} \tag{4.5}$$

$$ritotal_{it} = \alpha_{3i} + gdp\_pc_{it}\,\beta_{31} + tfc\_pc_{it}\beta_{32} + \sum_{k=-a}^{b}\Delta gdp\_pc_{it+k}\,\theta_{31} + \sum_{k=-a}^{b}\Delta\, tfc\_pc_{it+k}\theta_{32} + \omega_{it} \tag{4.6}$$

where gdp_pc_{it} is the per capita GDP of country i at time t, tfc_pc_{it} is the per capita energy consumption of country i at time t, $ritotal_{it}$ is the energy price index of country i at time t, β represents the long-run DOLS coefficients, θ represents vectors of coefficients on the lags and leads of the first-differenced explanatory variables, α_i denotes the country fixed effects and ε_{it}, ϵ_{it}, ω_{it} denote the various error terms. Maximum lag and lead lengths are shown by a and b, respectively.

Choosing the optimal lag and lead lengths involves the trade-off of increased precision vs. bias reduction. The standard method for making this choice for the DOLS estimation is based on minimising the Akaike information criterion. For equation 4.4, with *gdp_pc* as a dependent variable, this suggests six lags and leads. However, it turns out that the sixth lag makes no real difference to the estimation – after checking the robustness of the results to the lag and lead length, both the point estimates of the coefficients of interest and their standard errors barely change when going from five to six lags and leads. This suggests that accounting for more lags and leads does not result in bias reduction and therefore I proceed with five lags and leads. Similar analysis is carried out for all specifications.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable					
	gdp_pc	*tfc_pc*	*ritotal*	*gdp_pc*	*tfc_pc*	*ritotal*
gdp_pc	—	0.514***	0.507***	—	0.122**	0.064
		(0.000)	(0.000)		(0.045)	(0.575)
tfc_pc	0.680***	—	-0.709***	0.208	—	-0.675***
	(0.000)		(0.000)	(0.166)		(0.006)
ritotal	0.788***	-0.467***	—	0.290***	-0.233***	—
	(0.000)	(0.000)		(0.001)	(0.000)	
Lags and leads	5	1	2	5	1	2
Speed of adjustment	-0.055***	-0.070***	0.826***	-0.044*	-0.080**	0.412***
Subsample	28 OECD countries			G7 countries		
N	28			7		
Number of observations	700	924	868	175	231	217

Table 4.7. Dynamic OLS results

Notes: p-values in parentheses. *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

In addition to estimating the long-run coefficients for the whole sample, I also restrict the sample to G7 countries. The results are presented in table 4.7. Table 4.7 also includes the results from a two-step ECM, where in the first step the DOLS residuals have been collected, and in the second step the error correction model is estimated by OLS using cluster robust standard errors. The second step equations have the following form:

$$\Delta y_{it} = \gamma + \varphi dols\_residual\_y_{it-1} + \sum_{j=1} \Delta y_{it-j}\phi_j + \sum_{j=0} \Delta X_{it-j}\omega_j + \nu_{it}$$
(4.7)

and

$$dols\_residual\_y_{it-1} = y_{it} - X_{it}\,\beta - \alpha_i - \sum_{k=-a}^{b} \Delta X_{it+k}\theta \qquad (4.8)$$

where y is the dependent variable in the corresponding DOLS equation (4.4-4.6), which can be *gdp_pc, tfc_pc* or *ritotal*; X is a matrix that contains the other two variables; φ is the parameter of interest in the ECM as it indicates whether the dependent variable adjusts to the equilibrium and the magnitude represents the speed of adjustment toward the long-run relationship. We expect to find φ negative and significant if the dependent variable in the corresponding equation is error-correcting; ϕ and ω are each vectors of coefficients on the lagged first difference of the dependent variable and current and lagged differences of the regressors respectively, γ allows for a constant in the second step equation and ν_{it} is an error term.

For the whole sample, all long-run coefficients and speed of adjustment terms appear to be significantly different from zero at the 1% level. However, the speed of adjustment coefficient in the equation where energy prices are used as a dependent variable is positive and quite large in magnitude, which might suggest misspecification. In the other two equations, with *gdp_pc* and *tfc_pc* as dependent variables, speeds of adjustment suggest that both variables are error-correcting. The adjustment, however, happens very slowly.

The long-run coefficients suggest that a 1% increase in energy consumption leads to 0.7% increase in GDP and 1% increase in GDP results in 0.5% increase in energy consumption. The elasticity of energy consumption with respect to energy prices is negative. A negative and significant speed of adjustment can be interpreted as long-run causality, so based on the DOLS results there is bi-directional long-run causality between energy and economic growth.

The situation is different in the case of G7 countries. All speed of adjustment terms although similar in magnitudes and statistically significant, are less precisely estimated. The speed of adjustment term of energy prices is still positive, which as noted earlier might suggest that there are some problems with the model. However, the long-run coefficients are notably smaller in magnitude – a 1% increase in GDP predicts a 0.1% increase in total energy consumption, which is four times smaller than the effect for the whole sample. And total energy consumption doesn't appear to have any effect on GDP (see column 4), which in combination with the speed of adjustment that is only significantly different from zero at the 10% level, suggests that there is no long-run causality running from energy consumption to GDP. Also, energy consumption is less responsive to energy prices in G7 countries – the coefficient is about half of that for the whole sample, however the estimate of elasticity is still negative and significantly different from zero.

In conclusion, in G7 countries, the long-run relationship between the variables is much weaker and, unlike the whole sample, where bi-directional causality was found, the results suggest that causality goes from GDP to energy consumption, but not the other way around.

4.3.5 Long-run relationship (MG and PMG).

It has to be noted that while DOLS and the ECM based on the DOLS residuals produce consistent and efficient estimates, they don't allow for any heterogeneity in the cointegrating equation and impose same long-run and short-run coefficients as well as same error variances in all panel units, which is very restrictive for a macro panel. Only country-specific fixed effects can vary across panels.

In contrast to the majority of the papers in the field that use "micro panel" methods - GMM estimators, such as Arellano and Bond (1991) - to estimate the error correction model and the direction of causality, I opt for the Mean Group (MG) and Pooled Mean Group (PMG) estimators by Pesaran and Smith (1995) and Pesaran et al. (1999) respectively, which unlike panel GMM estimators, are more appropriate for the large-T and large-N setting. The main justification for Arellano and Bond type estimators is the correlation of the lagged dependent variable with the differenced error term that inflicts a bias on the fixed effects estimation of the differenced data. However, the bias is considered to be an issue only when T is small, which isn't the case in this analysis. Also, the micro panel methods impose homogenous dynamics on the panels, which is too restrictive in the macro setting.

The PMG estimator allows for different short-run coefficients in the different panel units (the different countries, in this case), while constraining the long-run relationship to be the same for all panel units. This is in contrast to DOLS, which only allows the constants to vary across panels. The MG estimator estimates a different equation for each panel and then averages the individual panel estimates. MG is most appropriate if the relationship varies sharply between countries and is always consistent. The PMG estimates are only consistent if the assumption of a common long-run relationship across panels is valid. The assumption can be tested using a Hausman test, and if the null of no difference between MG and PMG coefficients is not rejected, then preference is generally given to PMG as it is a more efficient estimator. If, on the other hand, the

Hausman test does reject the null, MG has to be chosen as it is the only consistent estimator in that case. While MG just takes unweighted averages of the estimates for each panel unit, the ECM estimated by PMG has the following form:

$$\Delta y_{it} = \varphi(y_{it-1} - \beta_1 x_{1it} - \beta_2 x_{2it} - \alpha_i) + \sum_{j=1}^{p} \Delta y_{i,t-j}\lambda_j + \sum_{j=0}^{q} \Delta x_{1i,t-j}\phi_{1j} + \sum_{j=0}^{q} \Delta x_{2i,t-j}\phi_{2j} + v_{it} \quad (4.9)$$

where y is the corresponding dependent variable (*gdp_pc, tfc_pc or ritotal*), β represents the long-run coefficients, φ denotes the error correction speed of adjustment coefficient, $x_1$and x_2 the other two variables that are treated as regressors in the corresponding equation, α_i once again denotes the country fixed effects, λ is a $(p \times 1)$ vector of parameters on the lagged differenced dependent variable, ϕ_1 and ϕ_2 are $((q+1) \times 1)$ vectors of parameters on the differenced regressors and lagged differenced regressors, v_{it} denotes the error term. PMG is estimated using maximum likelihood estimation.

The results of the estimations are presented in table 4.8. First, I consider columns (1) and (2), where all the observations in the sample are used and GDP per capita is used as a dependent variable. The speed of adjustment is similar in magnitude to the DOLS estimates and is also negative and significantly different from zero at the 1% level, which suggests slow adjustment of GDP to the equilibrium relationship. The magnitudes of the coefficients on *tfc_pc* vary widely depending on the method of estimation. The MG estimate is more than two times larger than the PMG estimate. However, the Hausman test statistic rejects only at the 10% significance level, but not at 5%, which suggests that the difference between the coefficients is on the borderline of statistical significance. This result is driven by the fact that the MG estimates have very large standard errors. The estimated coefficients of about 2.1 on energy consumption and of 0.85 on the energy price index have standard errors of 0.75 and 0.26 respectively, so despite being markedly less than half the size of the corresponding PMG estimates (0.95 and 0.29), the Hausman test cannot pick up the difference due to imprecision of the MG estimation. In fact, almost all MG estimates are associated with large standard errors, which is important for this analysis as it suggests that different panels produce very different point estimates when looked at individually, so restricting coefficients to be the same across panels will result in inconsistent estimation, even in cases where the Hausman test cannot reject the null of no systematic difference in the coefficients. This means we have to be very careful interpreting PMG as well as DOLS results, as both estimations impose restrictions on the coefficients that are likely not to hold.

The same can be said about estimates in columns (7) and (8), in which GDP is used as a dependent variable for the G7 subsample. The Hausman test doesn't reject the null hypothesis of no systematic difference between the coefficients, but we can see from the p-values that the standard errors are very large, so the failure by the Hausman test to reject the difference in the coefficients can be attributed to imprecision in the MG estimates rather than to the fact that the coefficients are actually close in magnitude. Does this mean that using panel estimation to evaluate the relationship between energy consumption and GDP is not necessarily the right thing to do, because different countries, even within the subset of G7 countries, appear to have very different long-run coefficients?

Not necessarily. Even though we have to be careful interpreting panel estimation results, and more so for estimators that do not allow heterogeneity in parameters, there are some results that appear to be robust. Consider results in columns (3) and (4), where all the observations are used and the total energy consumption is the dependent variable. The MG results are precise enough that the Hausman test strongly rejects the restriction of the same long-run coefficients even though the coefficients do not appear to be very different in magnitude. This implies that GDP and energy prices affect total energy consumption similarly in different countries. The same is true in columns (9) and (10), where the Hausman test rejects its null and suggests that MG is the appropriate estimator. Essentially, what is desired in a setting like this is to find MG estimates with relatively small standard errors, which would suggest that relationships across panels are similar and, only conditional on this first finding, the PMG estimation would pose a good alternative that can provide efficiency gains. However, if MG produces estimates with large standard errors, we are better of considering MG estimates even if the Hausman test suggests that MG and PMG estimates are not systematically different.

Consider again the overall results. The MG and PMG estimations suggest that for 28 OECD countries all the variables adjust towards the long-run relationship, and even energy prices have negative speed of adjustment coefficients, unlike in DOLS estimation, which reflects positively on the model. All the coefficients are significantly different from zero at the 1% level. However, the speeds with which the variables adjust are very different. Results in columns (1) and (3), suggest that energy consumption is more responsive to changes in the other two variables and adjusts to the equilibrium much faster than GDP per capita. The speed of adjustment coefficients are 0.096 for *gdp_pc* and 0.451 for *tfc_pc*, which suggests that energy consumption closes almost half of the disequilibrium gap in one year, while GDP only closes about 10% of the gap.

For G7 countries, GDP does not appear to adjust to the long-run relationship as the speed of adjustment term is not significantly different from zero at any conventional significance level, while energy consumption appears to be very responsive – the speed of adjustment coefficient is equal to 0.504 and is significant at the 1% level. Two things here are worth highlighting: 1. The speed of adjustment of GDP per capita (column 1) relative to the speed of adjustment of energy consumption (column 3) for the whole sample appears to be much slower. 2. The speed of adjustment coefficient in the equation with *gdp_pc* as a dependent variable for G7 countries (column 7) is insignificant. Taken together, these two results suggest that GDP, despite its participation in the equilibrium relationship, is probably not driven by energy consumption, or, if it is, the effect is quite small. This is also supported by the fact that long-run coefficients in column (7) are not significant at any conventional level.

At the same time, there is strong evidence that per capita income drives energy consumption as the speed of adjustment coefficients in columns (3) and (9) are significantly different from zero at the 1% level and large in magnitudes. They suggest that in both samples energy consumption responds strongly to the movements in income per capita and energy prices, and closes about 50% of the gap of a disequilibrium in one year. Also, the long-run coefficient suggests that a 1% increase in GDP per capita increases energy consumption by about 0.5% in the whole sample and by about 0.25% in G7 countries.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	MG	PMG	MG	PMG	MG	PMG	MG	PMG	MG	PMG	MG	PMG
	Dependent variables											
	gdp_pc	*gdp_pc*	*tfc_pc*	*tfc_pc*	*ritotal*	*ritotal*	*gdp_pc*	*gdp_pc*	*tfc_pc*	*tfc_pc*	*ritotal*	*ritotal*
gdp_pc			0.465***	0.791***	1.292**	1.120***			0.247**	0.805***	1.373***	0.771***
			(0.000)	(0.000)	(0.024)	(0.000)			(0.05)	(0.000)	(0.01)	(0.000)
tfc_pc	2.142***	0.950***			-1.920	-1.097***	1.291	1.011***			-0.043	-0.145
	(0.004)	(0.000)			(0.417)	(0.000)	(0.698)	(0.000)			(0.983)	(0.75)
ritotal	0.853***	0.292***	-0.500***	-0.687***			1.229	0.233**	-0.331***	-0.362***		
	(0.001)	(0.000)	(0.003)	(0.000)			(0.421)	(0.019)	(0.000)	(0.000)		
Speed of	-0.096***	-0.053***	-0.451***	-0.112***	-0.349***	-0.253***	-0.055	-0.058*	-0.504***	-0.108**	-0.215***	-0.191***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	0.103	0.051	(0.000)	0.048	(0.000)	(0.000)
Subsample	28 OECD countries						G7 countries					
Lags of first	2	2	2	1	2	2	1	1	2	2	2	2
№ of obs.	924	924	924	952	924	924	238	238	231	231	231	231
Hausman Test	5.07*		23.11***		0.12		2.21		23.15***		1.68	
Hausman Test	PMG		MG		PMG		PMG		MG		PMG	

Table 4.8. Mean Group and Pooled Mean Group estimation results

Notes: p-values in parentheses. *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

The last thing to consider here is whether there is any evidence of short-run causality among the variables. The short-run Granger causality tests can be performed by testing the null hypothesis that ϕ_{1j} or ϕ_{2j} jointly equal to zero in each equation. For example, if *gdp_pc* is chosen as a dependent variable and *tfc_pc* is the first regressor, failure to reject the null of ϕ_1 being equal to zero would suggest that energy consumption doesn't Granger-cause GDP in the short run.

	(1)	(2)	(3)	(4)	(5)	(6)
	MG	MG	MG	MG	MG	MG
Source of causation	Dependent variables					
	gdp_pc	*tfc_pc*	*ritotal*	*gdp_pc*	*tfc_pc*	*ritotal*
gdp_pc		61.48***	48.68***		38.39***	15.33***
		(0.000)	(0.000)		(0.000)	(0.002)
tfc_pc	31.16***		4.60	17.78***		63.92***
	(0.000)		(0.204)	(0.000)		(0.000)
ritotal	14.41***	14.63***		11.86***	9.02**	
	(0.002)	(0.002)		(0.003)	(0.03)	
Subsample	28 OECD countries			G7 countries		
Lags of FD	2	2	2	1	2	2
Number of observations	924	924	924	238	231	231

Table 4.9. Short-run causality tests (based on MG estimation results).

Notes: p-values in parentheses. *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

The results of these tests, presented in table 4.9, suggest that in the short run all three variables affect each other. One exception are energy prices, which do not appear to be affected by energy consumption in the whole sample.

4.3.6 Testing for cross-sectional dependence and CCEMG estimations.

In the previous section, I showed that the long-run relationship between income, energy consumption and energy prices is heterogeneous across panels, so I concluded that

PMG, which imposes the same long-run coefficients in all panels, is not an appropriate estimator, and MG should be used instead. However, both MG and PMG do not produce consistent results if errors exhibit cross-sectional dependence and the source of the common movements in the errors is correlated with the regressors.

I check for cross-sectional dependence of the errors using Pesaran's (2015) test. Under the null of weak cross-sectional dependence, the test statistic has standard normal distribution. The test is applied to the residuals from the MG estimations when the whole sample is used. The results are presented in table 4.10. As you can see, the test statistics are highly significant and the null of weak dependency is always rejected, regardless of which variable is used as a dependent variable. So I conclude that errors are strongly cross-sectionally dependent.

	Dependent variable		
Test statistic	*gdp_pc*	*tfc_pc*	*ritotal*
CD	25.97***	10.72***	41.74***
p-value	(0.000)	(0.000)	(0.000)

Table 4.10. Results of Pesaran's (2015) test for weak cross-sectional dependence (based on the MG residuals).

Notes: p-values in parentheses. *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

Failure to reject weak cross-sectional dependence of the errors suggests that there are common unobserved factors across panels, which are not captured by the model. These unobserved factors are likely to be correlated with the explanatory variables, which would bias the MG estimates of the long-run and short-run coefficients. I re-estimate the relationships using Pesaran's (2006) Common Correlated Effects Mean Group (CCEMG) estimator, which adds cross-sectional means of the variables in the model to proxy for unobserved common factors. The estimator is consistent in the presence of endogeneity induced by the unobserved factors.

Inclusion of the means increases the number of parameters to be estimated, so in order to preserve degrees of freedom I estimate a simpler ECM with only levels of variables and their first-differences. Following Mohammadi and Parvaresh (2014), I add means of levels and means of their first differences.

The estimated equation takes the following form:

$$\Delta y_{it} = \varphi_i y_{it-1} + \beta_{1i} x_{1it} + \beta_{2i} x_{2it} + \phi_{1i} \bar{y}_{it-1} + \phi_{2i} \bar{x}_{1it} + \phi_{3i} \bar{x}_{2it} + \alpha_i + \lambda_{1i} \Delta x_{1it} + \lambda_{2i} \Delta x_{2it} + \phi_{4i} \overline{\Delta y}_{it} + \phi_{5i} \overline{\Delta x}_{1it} + \phi_{6i} \overline{\Delta x}_{2it} + \nu_{it} \tag{4.10}$$

where y is the corresponding dependent variable (*gdp_pc, tfc_pc or ritotal*), β represents the long-run coefficients, φ denotes the error correction speed of adjustment coefficient, x_1 and x_2 the other two variables that are treated as regressors in the corresponding equation, α_i denotes the country fixed effects, ϕ represent coefficients on the means of levels of variables in their first differences, λ_1 and λ_2 are parameters on the differenced regressors, and ν_{it} denotes the error term.

The results of the estimation, presented in table 4.11, largely support earlier findings. It can be seen in column (1) that the effect of energy consumption on income varies significantly across countries, which manifests itself in large standard errors and a 95% confidence interval that includes zero. The effect of income change on energy consumption is much more precisely estimated (see column (2)). It is slightly larger in magnitude than the MG estimate (0.629 as opposed to 0.465), but appears to be consistent across panels and suggests that in most countries 1% increase in GDP per capita causes a 0.6% increase in energy consumption in the long run.

All speed of adjustment coefficients are negative and significantly different from zero at the 1% level, which suggests that all three variables adjust to the equilibrium relationship. However, the speed of adjustment coefficients are differ markedly across the various dependent variables, and they also differ across countries. The speed of adjustment estimate is largest in magnitude in the equation where *tfc_pc* is used as a dependent variable. In that case, the estimates suggest that movements in energy consumption close half of the disequilibrium gap in a year, which is very similar to the result obtained using the MG estimation. And while there was some suggestion noted earlier that the G7 differs from the broader OECD, I am unfortunately not able to estimate the relationship for the G7 countries separately, as consistency of CCEMG requires a larger number of panel units. Formally, the number of panel units should tend to infinity, but Pesaran (2006) tests the small sample properties of the estimator and finds them satisfactory, however these results apply when the number of panel units is at least 20, rather than only 7.

Variables	(1) CCEMG	(2) CCEMG	(3) CCEMG
	Dependent variables		
	gdp_pc	*tfc_pc*	*ritotal*
gdp_pc		0.629***	-0.657
		(0.002)	(0.402)
tfc_pc	0.407		3.275
	(0.486)		(0.309)
ritotal	-0.342	-0.180*	
	(0.363)	(0.092)	
Speed of adjustment	-0.249***	-0.528***	-0.490***
	(0.000)	(0.000)	(0.000)
d.gdp_pc		0.12	0.026
		(0.159)	0.873
d.tfc	0.138***		
	(0.000)		
d.ritotal	0.007	0.015	-0.194*
	(0.665)	(0.680)	(-0.053)
Subsample	28 OECD countries		
Lags of FD	0	0	0
Number of observations	980	980	980

Table 4.11. Common Correlated Effects Mean Group estimation results

Notes: p-values in parentheses. *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

I test the CCEMG residuals for weak cross-sectional dependence using Pesaran's (2015) test. Despite the fact that the null of weak cross-sectional dependence is still rejected for the equation with *tfc_pc* as a dependent variable, all the test statistics are much smaller than those calculated based on the MG residuals, which indicates that cross-sectional means captured a large fraction of the common unobserved trends.

Test statistic	Dependent variable		
	gdp_pc	*tfc_pc*	*ritotal*
CD	-1.63	-2.38**	-1.54
p-value	(0.103)	(0.02)	(0.123)

Table 4.12. Results of Pesaran's (2015) test for weak cross-sectional dependence (based on the CCEMG residuals).

Notes: p-values in parentheses. *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

One caveat here is that the ECM parameter estimates might be sensitive to the number of lags of the first-differenced variables in the model. However, as the number of times periods is not very large, it is not possible to incorporate more lags without compromising the reliability of estimations – there simply wouldn't be enough degrees of freedom left to obtain any meaningful results if more lags are added to the model. For the same reason, I do not test the long-run relationship for the presence of structural breaks. I've argued in this chapter that due to parameter heterogeneity across countries, the mean group estimators should be chosen. But since they are not as efficient as some other estimators (e.g. DOLS or PMG), I am unable to test some parameter-intense specifications, such as structural breaks or elaborate lag structures.

4.4 Conclusion

4.4.1 Econometric results

The contribution of this chapter has been to attempt to clarify the nature of the relationship between energy consumption and economic growth using newer econometric techniques than had previously been used in the (somewhat conflicting) literature. Following Bai and Ng (2004, 2011), I start by decomposing the variables (GDP per capita, energy consumption per capita and energy prices) into common factors and idiosyncratic factors by the means of PCA and testing common factors and idiosyncratic factors for stationarity separately, which controls for potential cross-sectional dependences in the variables. The results of the tests suggest that both common factors and idiosyncratic factors are non-stationary, which, contrary to Belke et al. (2011), suggests that the long-run relationship between the variables is not purely driven by common international trends.

Next, I test for the existence of the long-run cointegrating relationship between the variables using the Westerlund cointegration tests with bootstrapped standard errors, which are robust to the presence of common factors across panel units. The results are conflicting, with two test statistics supporting presence of a long-run relationship and two which do not support the presence of this relationship.

I continue by estimating the long-run relationship and the error-correction model using panel DOLS estimation, which suggests there is bi-directional long-run causality between energy consumption and economic growth in the sample of 28 OECD countries, but among G7 countries, the long-run relationship between the variables is much weaker and the causality goes from GDP per capita to energy consumption, but not the other way around, which supports the conservation hypothesis.

In order to test and account for potential heterogeneity of the parameters across panel units, I use the MG and PMG estimators to estimate the relationship. The MG estimates have large standard errors in the equations that test if income is affected by energy consumption, which suggests that the relationship is heterogeneous across panels. However, GDP per capita appears to affect energy consumption similarly across countries and the effect is much stronger. There is some evidence that income is driven by energy consumption in the long run in the whole sample, but the effect is weak. For G7 countries, no long-run causality from energy consumption to GDP was found. At the same time, the results of the estimations suggest that energy consumption is driven by economic growth for the whole sample and for the subset of G7 countries in the long run, and in both samples, the degree of responsiveness of energy consumption is quite high. In the short run, energy consumption and GDP exhibit mutual causality. The elasticity of energy consumption with respect to energy prices is negative for the whole sample and for G7 countries, but the magnitude for 28 OECD countries is double of that for G7 countries.

Lastly, I test the MG residuals for weak cross-sectional dependence and reject it in favour of strong dependence, which indicates potential bias in the MG estimations. I re-estimate the relationships using the CCEMG estimator, which is robust to strong cross-sectional dependence in the errors. The results, although different in magnitude from the MG estimation results, lead to similar conclusions – the effect of energy consumption on GDP varies significantly across countries, while the effect of GDP per capita on energy consumption is much more homogeneous; the direction of causality is bi-directional with energy consumption responding faster to disequilibrium.

The results obtained in this analysis suggest that the conclusions are sensitive to the econometric methods used and failure to allow for heterogeneity in parameters might be distorting the magnitudes of the effects. However, despite strong evidence in favour of heterogeneity of the effect of energy consumption on economic growth, all econometric methods employed in the analysis point at the same direction of causality – for the whole sample the causality is bi-directional, but for the G7 countries, causality runs one way – from economic growth to energy consumption. Why might the results be different for the G7 as compared to the rest of the OECD? One possibility is that for a country like New Zealand, say, energy prices really are exogenous and when they go up, the economy suffers. But for a larger economy like the United States, energy prices will in a greater sense be determined by the business cycle – energy prices rise when the American economy is doing well. This complicates the causal link between energy consumption and GDP and may result in the finding highlighted above.

Thus the current study suggests that the results are not only sensitive to the methods used, but also to the way countries are grouped. Panel econometric methods are useful in analysing the relationship between energy consumption and economic growth, however, many studies neglect that fact that most panel estimators produce meaningful results only if the imposed restriction of homogeneity of parameters is valid. Based on the results of my analysis, this restriction is not likely to hold. It will be particularly beneficial for the future research in this field to either group countries based on the similarity of the long-run relationship relying on past findings, or conduct individual country-specific analysis first, and then apply appropriate panel methods to make use of increased precision of estimation.

4.4.2 Policy Implications

World primary energy use is projected to grow by 1.3% per year between 2015 and 2035, but almost all of this growth is expected to come from the developing world, with China and India alone accounting for over half the increase; energy demand in the European Union is actually predicted to fall by 0.4% per year over the same period (BP (2017)). Given the strong historical relationship between energy use and GDP growth, should GDP growth forecasts be revised downwards when energy use forecasts are revised downwards? Is this something that European countries should worry about?

Taken at face value, the results of the current study do not suggest that there is any reason for panic on these grounds. The main finding was that the correlation between energy usage and GDP was driven primarily by a one-directional relationship where higher levels of economic growth were (Granger) causing higher levels of energy usage. There was no strong evidence for higher (or lower) energy usage leading to higher (or lower) economic growth.

Given the increasing shift of global energy use towards the developing world, it would be useful if the current results could say something about the relationship between energy and GDP in the developing world also. Unfortunately, however, the sample used in the present study consisted entirely of OECD countries, so this would push up against the limits of external validity. The sample used here begins in 1978, when the poorest sample countries (measured in 2005 US$) were: South Korea (GDP per capita: $4,650), Mexico ($9,130), and Poland ($9,380). The GDP per capita of China overtook South Korea's 1978 value only recently (around 2011), whereas India remains below half of the South Korean 1978 value of GDP per capita even in 2016 (World Bank). It would thus be unwise to infer much about a set of developing countries based on data where even the poorest sample country (nevermind the average) was so much wealthier. Doubly so given the well-known greater importance of more energy-intensive activities such as manufacturing or construction to developing countries like China. Further work would be needed in this area before any firm statements could be made about whether (for example) China's efforts to meet her commitments to the Paris Climate Accord will noticeably impair Chinese economic growth.

Concluding chapter.

This conclusion collects some practical recommendaitons emerging from the three substantial chapters of this thesis.

From the second chapter (*Should we pre-test instrumental variables? A Monte Carlo study.*), the strong recommendation to IV practitioners is to rely on prior information about the strength of instruments and the degree of endogeneity to inform the decision about whether to pre-test results. If the instrument is suspected to be weak and the degree of endogeneity is suspected to be high, then a high first-stage F-statistic is likely to correspond to a bias in the IV estimation. Researchers should instead consider statistics which are robust to weak instruments, such as AR (1949) and Kleibergen's (2002) Lagrange Multiplier statistic. These statistics have distributions that do not depend on the value of the concentration parameter. A better alternative is to use split-sample IV, which eliminates pre-test bias. But this requires a large sample, and if the correlation between the instrument and the regressor is low enough, then it will significantly impair the estimation of coefficients, so this is yet another case where researchers must focus on the bias-variance tradeoff. However all of this is predicated on the joint assumption of weak instruments and high endogeneity: if the degree of endogeneity is sufficiently low, choosing instruments that exceed a threshold value is useful – the MSE is lower for the samples conditioned on the high F-statistic, and the corresponding test size distortions are also low.

From the third chapter (*Oil Rents and the Real Exchange Rate.*), the recommendations would be directed more at policy makers or perhaps development economists rather than econometric practicioners. In that chapter, we used proprietary data to check whether we could find evidence for the Balassa-Samuelson effect (the notion that productivity changes in the tradable sector lead to real appreciations of the currency) within the oil sector in major oil exporting countries. While we did find evidence that productivity in the oil sector has an effect on the real exchange rate in some countries, there are two major caveats to the finding: (i) where there was an effect, it was small in magnitude compared to the effect of oil prices on the exchange rate, and (ii) there was no significant effect in OPEC, which contains the most significant oil exporting countries. In practical terms, these caveats suggest that whatever difficulties non-oil industries face within major oil exporting countries (e.g. if there is a 'resource curse' for oil), these difficulties do *not* seem to stem from the oil sector causing an overvalued exchange rate, but from other (institutional?) factors.

From the fourth chapter (*How Costly is Conservation? The International Energy-GDP Relationship Re-examined*), the practical takeaways are about the relationship between energy and GDP. The main conclusion was that (at least for the G7 countries) causality runs from GDP to energy usage, but not vice versa. If correct, this finding would come as some relief to policymakers who are targeting lower energy usage and worried about the downside potential for economic growth. A more nuanced takeaway for those doing applied work in this area is that the degree of heterogeneity in the Energy→GDP channel appears to be far greater than in the GDP→Energy channel. This may be because economies differ substantially in (e.g.) the degree to which they rely on energy-intensive industries such as manufacturing as the basis for their economies, but their consumers all increase energy usage in response to higher incomes in more-or-less similar ways. In practical terms, this means that it may not be wise to try to estimate the Energy→GDP effect using panel data, since there may not be a 'typical' effect which is meaningfully stable across countries or time.

So, a few years' work in nutshell: use prior information sensibly when estimating, and don't worry so much about Dutch disease or the negative economic consequences of scaling back on energy usage.

Appendix.

Table A.1 Country-by-country ADF test results (lrer).

ISO-code	0 lags		1 lags		2 lags		0 lags		1 lags		2 lags	
	without trend						with trend					
	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n
AGO	0.676	23	0.658	22	0.709	21	0.921	23	0.923	22	0.952	21
ARE	0.902	23	0.826	22	0.873	21	0.595	23	0.314	22	0.417	21
ARG	0.603	17	0.663	16	0.680	15	0.632	17	0.669	16	0.676	15
AUS	0.181	39	0.016	38	0.045	37	0.265	39	0.010	38	0.012	37
AZE	0.987	11	0.539	10	0.854	9	0.331	11	0.243	10	0.091	9
BRA	0.655	16	0.347	15	0.032	14	0.911	16	0.755	15	0.276	14
BRN	0.007	39	0.000	38	0.000	37	0.222	39	0.008	38	0.001	37
CAN	0.677	28	0.060	27	0.176	26	0.934	28	0.179	27	0.475	26
CHN	0.994	12	0.932	11	0.948	10	0.968	12	0.997	11	1.000	10
COG	0.371	35	0.172	34	0.218	33	0.804	35	0.595	34	0.578	33
COL	0.684	23	0.076	22	0.075	21	0.863	23	0.206	22	0.194	21
DNK	0.424	36	0.048	35	0.184	34	0.674	36	0.117	35	0.319	34
DZA	0.470	44	0.477	43	0.498	42	0.688	44	0.729	43	0.663	42
ECU	0.635	12	0.350	11	0.773	10	0.375	12	0.000	11	0.057	10
EGY	0.463	44	0.159	43	0.309	42	0.549	44	0.053	43	0.131	42
GAB	0.234	44	0.247	43	0.183	42	0.485	44	0.519	43	0.445	42
GBR	0.160	40	0.014	39	0.041	38	0.373	40	0.005	39	0.008	38
GNQ	0.906	17	0.770	16	0.246	15	0.964	17	0.944	16	0.511	15
IDN	0.425	41	0.463	40	0.567	39	0.566	41	0.624	40	0.692	39
IND	0.851	10	0.706	9	0.844	8	0.517	10	0.299	9	0.633	8
IRQ	0.727	39	0.715	38	0.612	37	0.827	39	0.797	38	0.604	37
ITA	0.618	44	0.371	43	0.480	42	0.346	44	0.063	43	0.108	42
KAZ	0.847	14	0.767	13	0.762	12	0.899	14	0.442	13	0.315	12
LBY	0.763	23	0.687	22	0.579	21	0.566	23	0.396	22	0.093	21
MEX	0.542	15	0.032	14	0.356	13	0.073	15	0.983	14	0.994	13
MYS	0.715	36	0.624	35	0.743	34	0.522	36	0.089	35	0.356	34
NGA	0.338	35	0.163	34	0.074	33	0.497	35	0.220	34	0.055	33
NOR	0.608	31	0.152	30	0.236	29	0.589	31	0.122	30	0.113	29
OMN	0.000	35	0.003	34	0.021	33	0.004	35	0.005	34	0.012	33
PER	0.338	29	0.550	28	0.466	27	0.475	29	0.714	28	0.606	27
QAT	0.792	23	0.956	22	0.292	21	0.734	23	0.977	22	0.794	21
ROM	0.796	20	0.724	19	0.594	18	0.039	20	0.233	19	0.296	18
RUS	0.051	19	0.030	18	0.140	17	0.028	19	0.000	18	0.366	17
SAU	0.857	23	0.769	22	0.958	21	0.593	23	0.232	22	0.599	21
SDN	0.871	14	0.905	13	0.910	12	0.001	14	0.744	13	0.012	12
SYR	0.525	35	0.395	34	0.593	33	0.367	35	0.088	34	0.217	33
THA	0.661	24	0.267	23	0.450	22	0.915	24	0.576	23	0.749	22
TTO	0.410	34	0.186	33	0.225	32	0.675	34	0.486	33	0.507	32
TUN	0.186	42	0.146	41	0.066	40	0.314	42	0.231	41	0.069	40
VEN	0.948	20	0.906	19	0.879	18	0.882	20	0.813	19	0.569	18
VNM	0.046	19	0.003	18	0.075	17	0.546	19	0.003	18	0.022	17
YEM	0.659	20	0.288	19	0.330	18	0.942	20	0.506	19	0.787	18

Table A.2. Country-by-country ADF test results (loilrents_pc).

ISO-code	0 lags		1 lags		2 lags		0 lags		1 lags		2 lags	
	without trend						with trend					
	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n
AGO	0.567	23	0.612	22	0.853	21	0.603	23	0.483	22	0.815	21
ARE	0.297	23	0.209	22	0.477	21	0.543	23	0.380	22	0.769	21
ARG	0.647	17	0.422	16	0.613	15	0.038	17	0.001	16	0.180	15
AUS	0.000	39	0.000	38	0.034	37	0.000	39	0.000	38	0.027	37
AZE	0.481	11	0.424	10	0.227	9	0.805	11	0.596	10	0.167	9
BRA	0.038	16	0.193	15	0.017	14	0.060	16	0.397	15	0.002	14
BRN	0.003	39	0.000	38	0.000	37	0.108	39	0.010	38	0.000	37
CAN	0.018	28	0.172	27	0.290	26	0.053	28	0.302	27	0.402	26
CHN	0.669	12	0.346	11	0.739	10	0.034	12	0.422	11	0.879	10
COG	0.000	35	0.012	34	0.002	33	0.000	35	0.038	34	0.028	33
COL	0.184	23	0.574	22	0.705	21	0.306	23	0.530	22	0.777	21
DNK	0.313	36	0.306	35	0.144	34	0.095	36	0.074	35	0.002	34
DZA	0.000	44	0.171	43	0.305	42	0.000	44	0.513	43	0.674	42
ECU	0.192	12	0.043	11	0.642	10	0.065	12	0.015	11	0.049	10
EGY	0.402	44	0.194	43	0.151	42	0.865	44	0.711	43	0.708	42
GAB	0.072	44	0.133	43	0.111	42	0.145	44	0.428	43	0.411	42
GBR	0.334	40	0.233	39	0.226	38	0.637	40	0.466	39	0.476	38
GNQ	0.392	17	0.527	16	0.768	15	0.183	17	0.076	16	0.077	15
IDN	0.156	41	0.182	40	0.180	39	0.455	41	0.456	40	0.404	39
IND	0.100	10	0.753	9	0.767	8	0.155	10	0.855	9	0.419	8
IRQ	0.251	39	0.345	38	0.282	37	0.477	39	0.596	38	0.438	37
ITA	0.526	44	0.174	43	0.156	42	0.877	44	0.684	43	0.628	42
KAZ	0.764	14	0.609	13	0.870	12	0.709	14	0.080	13	0.171	12
LBY	0.540	23	0.436	22	0.723	21	0.645	23	0.470	22	0.851	21
MEX	0.555	15	0.577	14	0.852	13	0.456	15	0.328	14	0.506	13
MYS	0.001	36	0.095	35	0.036	34	0.021	36	0.347	35	0.207	34
NGA	0.255	35	0.355	34	0.482	33	0.557	35	0.603	34	0.856	33
NOR	0.156	31	0.316	30	0.797	29	0.077	31	0.063	30	0.341	29
OMN	0.060	35	0.113	34	0.359	33	0.218	35	0.349	34	0.732	33
PER	0.201	29	0.215	28	0.240	27	0.206	29	0.321	28	0.478	27
QAT	0.534	23	0.658	22	0.861	21	0.485	23	0.595	22	0.885	21
ROM	0.912	20	0.899	19	0.909	18	0.695	20	0.260	19	0.527	18
RUS	0.701	19	0.627	18	0.864	17	0.573	19	0.543	18	0.685	17
SAU	0.600	23	0.487	22	0.792	21	0.593	23	0.500	22	0.866	21
SDN	0.680	14	0.689	13	0.742	12	0.451	14	0.352	13	0.633	12
SYR	0.063	35	0.084	34	0.181	33	0.229	35	0.279	34	0.480	33
THA	0.901	24	0.927	23	0.986	22	0.240	24	0.745	23	0.814	22
TTO	0.042	34	0.009	33	0.340	32	0.149	34	0.032	33	0.712	32
TUN	0.016	42	0.004	41	0.172	40	0.067	42	0.017	41	0.406	40
VEN	0.086	20	0.277	19	0.635	18	0.138	20	0.303	19	0.671	18
VNM	0.344	19	0.491	18	0.734	17	0.024	19	0.111	18	0.661	17
YEM	0.005	20	0.120	19	0.307	18	0.003	20	0.047	19	0.050	18

Table A.3. Country-by-country ADF test results (lrgdpch).

ISO-code	0 lags		1 lags		2 lags		0 lags		1 lags		2 lags	
	without trend						with trend					
	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n
AGO	0.990	23	0.994	22	0.990	21	0.935	23	0.960	22	0.896	21
ARE	0.123	23	0.064	22	0.319	21	0.063	23	0.004	22	0.352	21
ARG	0.958	17	0.808	16	0.868	15	0.959	17	0.622	16	0.642	15
AUS	0.994	39	0.993	38	0.990	37	0.147	39	0.051	38	0.021	37
AZE	0.987	11	0.960	10	0.973	9	0.882	11	0.025	10	0.677	9
BRA	0.872	16	0.931	15	0.986	14	0.817	16	0.718	15	0.843	14
BRN	0.800	39	0.603	38	0.754	37	0.462	39	0.197	38	0.416	37
CAN	0.865	28	0.210	27	0.709	26	0.757	28	0.347	27	0.648	26
CHN	0.999	12	0.964	11	0.998	10	0.791	12	0.458	11	0.983	10
COG	0.494	35	0.096	34	0.033	33	0.846	35	0.451	34	0.187	33
COL	0.967	23	0.928	22	0.905	21	0.942	23	0.705	22	0.421	21
DNK	0.684	36	0.440	35	0.291	34	0.914	36	0.773	35	0.884	34
DZA	0.549	44	0.224	43	0.346	42	0.488	44	0.378	43	0.513	42
ECU	0.953	12	0.952	11	0.860	10	0.215	12	0.000	11	0.272	10
EGY	0.985	44	0.958	43	0.948	42	0.541	44	0.555	43	0.627	42
GAB	0.154	44	0.078	43	0.057	42	0.564	44	0.337	43	0.257	42
GBR	0.774	40	0.668	39	0.739	38	0.901	40	0.373	39	0.593	38
GNQ	0.602	17	0.437	16	0.003	15	0.986	17	0.889	16	0.985	15
IDN	0.090	41	0.310	40	0.320	39	0.479	41	0.281	40	0.367	39
IND	0.996	10	0.990	9	0.596	8	0.243	10	0.437	9	0.625	8
IRQ	0.161	39	0.164	38	0.246	37	0.431	39	0.435	38	0.547	37
ITA	0.000	44	0.009	43	0.009	42	0.980	44	0.982	43	0.991	42
KAZ	0.992	14	0.808	13	0.531	12	0.314	14	0.176	13	0.020	12
LBY	0.247	23	0.055	22	0.035	21	0.353	23	0.247	22	0.167	21
MEX	0.706	15	0.027	14	0.304	13	0.372	15	0.785	14	0.853	13
MYS	0.604	36	0.716	35	0.429	34	0.872	36	0.775	35	0.692	34
NGA	0.589	35	0.454	34	0.406	33	0.948	35	0.894	34	0.976	33
NOR	0.459	31	0.624	30	0.702	29	0.977	31	0.516	30	0.861	29
OMN	0.019	35	0.515	34	0.422	33	0.000	35	0.071	34	0.069	33
PER	0.947	29	0.575	28	0.825	27	0.938	29	0.515	28	0.791	27
QAT	0.998	23	0.995	22	0.992	21	0.948	23	0.961	22	0.910	21
ROM	0.986	20	0.664	19	0.833	18	0.130	20	0.008	19	0.699	18
RUS	0.939	19	0.401	18	0.642	17	0.107	19	0.014	18	0.179	17
SAU	0.695	23	0.298	22	0.404	21	0.488	23	0.447	22	0.231	21
SDN	0.988	14	0.966	13	0.979	12	0.200	14	0.460	13	0.231	12
SYR	0.369	35	0.739	34	0.862	33	0.273	35	0.573	34	0.673	33
THA	0.126	24	0.129	23	0.141	22	0.848	24	0.346	23	0.219	22
TTO	0.990	34	0.954	33	0.884	32	0.995	34	0.988	33	0.974	32
TUN	0.275	42	0.251	41	0.241	40	0.264	42	0.368	41	0.176	40
VEN	0.265	20	0.064	19	0.343	18	0.591	20	0.228	19	0.668	18
VNM	0.944	19	0.894	18	0.971	17	0.677	19	0.081	18	0.469	17
YEM	0.484	20	0.769	19	0.269	18	0.859	20	0.834	19	0.816	18

Table A.4 Country-by-country ADF test results (openc).

ISO-code	0 lags		1 lags		2 lags		0 lags		1 lags		2 lags	
	without trend						with trend					
	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n
AGO	0.175	23	0.373	22	0.318	21	0.581	23	0.913	22	0.877	21
ARE	0.398	23	0.237	22	0.205	21	0.760	23	0.564	22	0.530	21
ARG	0.632	17	0.697	16	0.649	15	0.737	17	0.862	16	0.684	15
AUS	0.588	39	0.591	38	0.616	37	0.038	39	0.062	38	0.118	37
AZE	0.781	11	0.176	10	0.238	9	0.992	11	0.973	10	0.911	9
BRA	0.731	16	0.669	15	0.261	14	0.965	16	0.976	15	0.888	14
BRN	0.049	39	0.042	38	0.221	37	0.168	39	0.149	38	0.524	37
CAN	0.749	28	0.437	27	0.401	26	0.996	28	0.993	27	0.986	26
CHN	0.664	12	0.201	11	0.409	10	1.000	12	0.996	11	0.996	10
COG	0.548	35	0.490	34	0.720	33	0.368	35	0.382	34	0.678	33
COL	0.690	23	0.745	22	0.668	21	0.192	23	0.720	22	0.835	21
DNK	0.720	36	0.757	35	0.648	34	0.493	36	0.255	35	0.348	34
DZA	0.505	44	0.152	43	0.345	42	0.763	44	0.350	43	0.661	42
ECU	0.325	12	0.198	11	0.422	10	0.475	12	0.362	11	0.377	10
EGY	0.312	44	0.178	43	0.123	42	0.524	44	0.362	43	0.305	42
GAB	0.005	44	0.037	43	0.101	42	0.001	44	0.004	43	0.076	42
GBR	0.055	40	0.074	39	0.048	38	0.087	40	0.115	39	0.094	38
GNQ	0.322	17	0.396	16	0.120	15	0.593	17	0.499	16	0.108	15
IDN	0.058	41	0.184	40	0.279	39	0.019	41	0.286	40	0.539	39
IND	0.729	10	0.976	9	0.872	8	0.067	10	0.532	9	0.392	8
IRQ	0.232	39	0.266	38	0.163	37	0.543	39	0.588	38	0.435	37
ITA	0.306	44	0.320	43	0.307	42	0.517	44	0.511	43	0.509	42
KAZ	0.427	14	0.421	13	0.437	12	0.870	14	0.938	13	0.963	12
LBY	0.885	23	0.830	22	0.719	21	0.846	23	0.811	22	0.756	21
MEX	0.000	15	0.287	14	0.448	13	0.000	15	0.555	14	0.757	13
MYS	0.545	36	0.710	35	0.589	34	0.988	36	0.900	35	0.968	34
NGA	0.154	35	0.479	34	0.445	33	0.068	35	0.405	34	0.313	33
NOR	0.139	31	0.014	30	0.219	29	0.179	31	0.011	30	0.361	29
OMN	0.131	35	0.013	34	0.889	33	0.296	35	0.001	34	0.714	33
PER	0.241	29	0.357	28	0.311	27	0.387	29	0.583	28	0.538	27
QAT	0.282	23	0.280	22	0.123	21	0.388	23	0.148	22	0.200	21
ROM	0.301	20	0.344	19	0.111	18	0.522	20	0.907	19	0.960	18
RUS	0.000	19	0.000	18	0.007	17	0.003	19	0.000	18	0.104	17
SAU	0.912	23	0.755	22	0.833	21	0.935	23	0.773	22	0.917	21
SDN	0.173	14	0.263	13	0.497	12	0.624	14	0.939	13	0.898	12
SYR	0.467	35	0.670	34	0.617	33	0.365	35	0.632	34	0.604	33
THA	0.541	24	0.502	23	0.567	22	0.430	24	0.743	23	0.804	22
TTO	0.115	34	0.232	33	0.189	32	0.260	34	0.298	33	0.271	32
TUN	0.335	42	0.133	41	0.198	40	0.202	42	0.026	41	0.100	40
VEN	0.533	20	0.389	19	0.355	18	0.739	20	0.663	19	0.601	18
VNM	0.745	19	0.823	18	0.989	17	0.139	19	0.065	18	0.375	17
YEM	0.330	20	0.106	19	0.340	18	0.940	20	0.916	19	0.972	18

Table A.5 Westerlund (2007) cointegration tests for lrer with lbrent and lrgdpch.

Dependent variable: *lrer*	**Test statistic**		**(1)**	**(2)**	**(3)**
lbrent		z-value	0.553	-5.751***	-5.773***
	Pt	p-value	0.710	0.000	0.000
lbrent		z-value	1.216	-5.133***	-4.601***
	Pa	p-value	0.888	0.000	0.000
lrgdpch		z-value	-2.031	2.091	0.839
	Pt	p-value	0.021**	0.982	0.799
lrgdpch		z-value	-0.487	3.159	1.676
	Pa	p-value	0.313	0.999	0.953
Subsample			OPEC	World-OPEC	D10-OPEC
N			10	32	16
Lags			1	1	1

Note: *, **, *** denote significance at 10%, 5%, and 1% level, respectively.

Figure A.1. GDP per capita PPP in constant 2005 USD by country

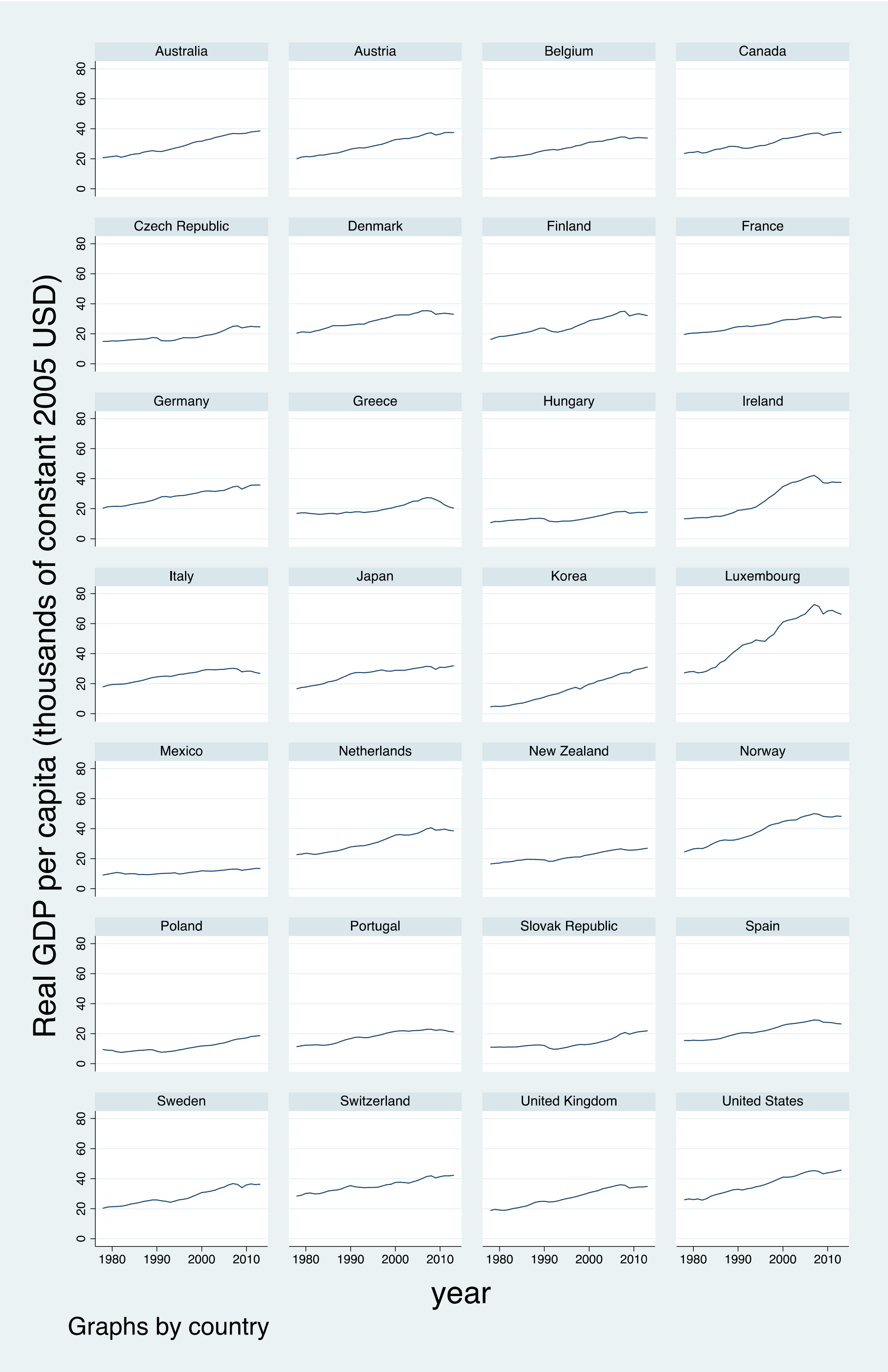

Figure A.2. Total final energy consumption per capita (in tonnes of oil equivalent) by country

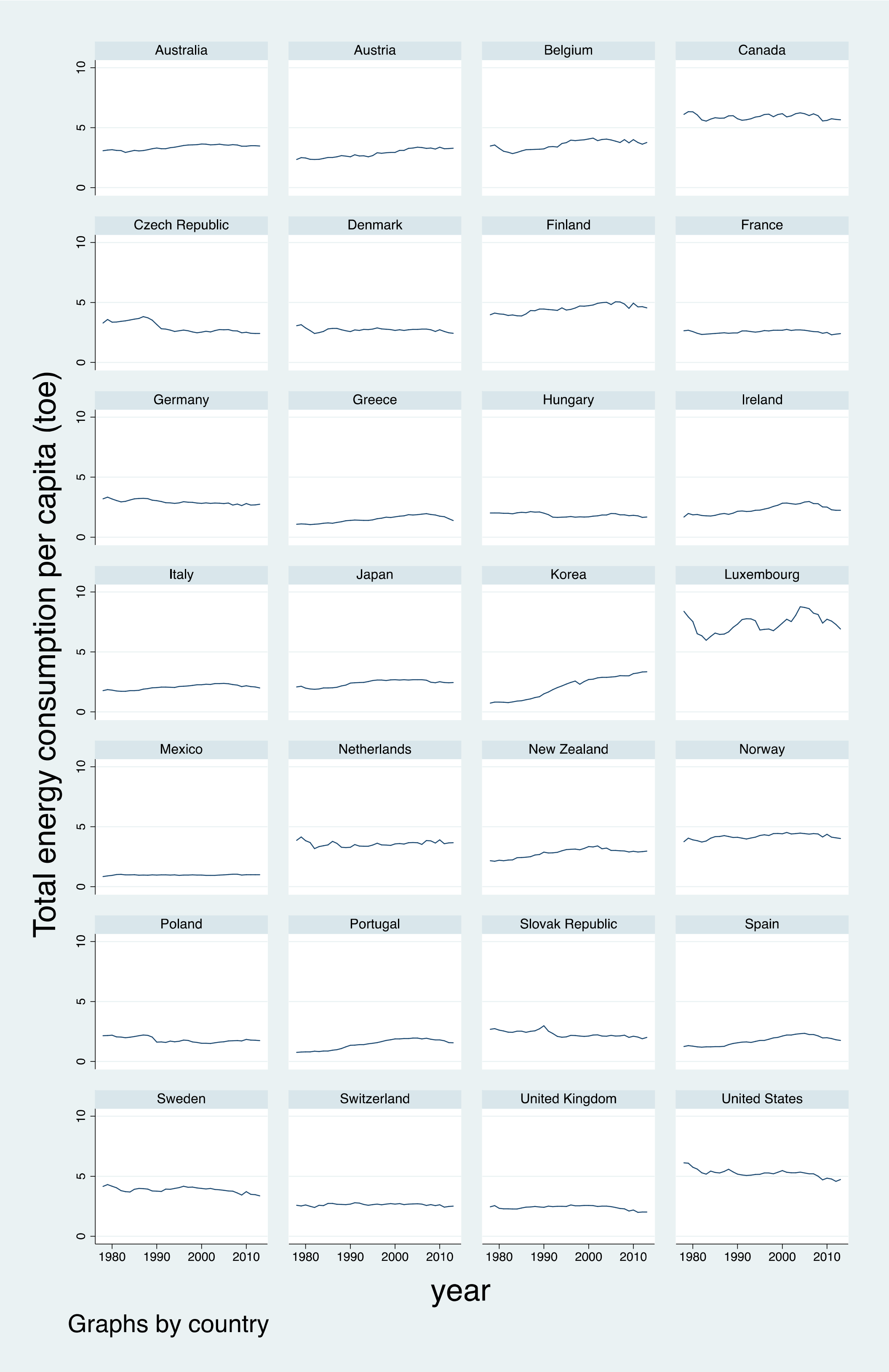

Figure A.3. Real end-use energy price index for industry and households

Bibliography

Akarca, Ali T., and Thomas Veach Long. "On the relationship between energy and GNP: a reexamination." *The Journal of Energy and Development* (1980): 326-331.

Al-Iriani, Mahmoud A. "Energy–GDP relationship revisited: an example from GCC countries using panel causality." *Energy policy* 34.17 (2006): 3342-3350.

Amin, Z. A., and M. I. T. El-Sakka. "Determining Real Exchange Rate Fluctuations in the Oil-Based GCC Economies." *Asian Economic and Financial Review* 6.7 (2016): 374.

Amiti, Mary, and Jozef Konings. "Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia." *The American Economic Review* 97.5 (2007): 1611-1638.

Apergis, Nicholas, and James E. Payne. "Energy consumption and economic growth: evidence from the Commonwealth of Independent States." *Energy Economics* 31.5 (2009): 641-647.

Anderson, Theodore W., and Herman Rubin. "Estimation of the parameters of a single equation in a complete system of stochastic equations." *The Annals of Mathematical Statistics* (1949): 46-63.

Angrist, Joshua D., and Alan B. Krueger. "Does Compulsory School Attendance Affect Schooling and Earnings?." *The Quarterly Journal of Economics* 106.4 (1991): 979-1014.

Angrist, Joshua D., and Alan B. Krueger. "Split-sample instrumental variables estimates of the return to schooling." *Journal of Business & Economic Statistics* 13.2 (1995): 225-235.

Apergis, Nicholas, and James E. Payne. "The emissions, energy consumption, and growth nexus: evidence from the commonwealth of independent states." *Energy Policy* 38.1 (2010a): 650-655.

Apergis, Nicholas, and James E. Payne. "Energy consumption and growth in South America: Evidence from a panel error correction model." *Energy Economics* 32.6 (2010b): 1421-1426.

Apergis, Nicholas, and James E. Payne. "Renewable and non-renewable energy consumption-growth nexus: Evidence from a panel error correction model." *Energy Economics* 34.3 (2012): 733-738.

Arellano, Manuel, and Stephen Bond. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *The review of economic studies* 58.2 (1991): 277-297.

Asafu-Adjaye, John. "The relationship between energy consumption, energy prices and economic growth: time series evidence from Asian developing countries." *Energy economics* 22.6 (2000): 615-625.

Asea, Patrick K., and Enrique G. Mendoza. "The Balassa-Samuelson Model: A General-Equilibrium Appraisal." *Review of International Economics* 2.3 (1994): 244-267.

Atkinson, Giles, and Kirk Hamilton. "Savings, growth and the resource curse hypothesis." *World Development* 31.11 (2003): 1793-1807.

Aziz, Mukhriz Izraf Azman, and Abu Bakar. "Oil price & exchange rate: A comparative study between net oil exporting and net oil importing countries." *ESDS International Annual Conference, London*. 2009.

Bai, Jushan, and Serena Ng. "A PANIC attack on unit roots and cointegration." *Econometrica* 72.4 (2004): 1127-1177.

Bai, Jushan, and Serena Ng. "Panel unit root tests with cross-section dependence: a further investigation." *Econometric Theory* 26.04 (2010): 1088-1114.

Balassa, Bela. "The purchasing-power parity doctrine: a reappraisal." *Journal of political Economy* 72.6 (1964): 584-596.

Bahmani-Oskooee, Mohsen, and Hyun-Jae Rhee. "Time-Series Support for Balassa's Productivity-Bias Hypothesis: Evidence from Korea." *Review of International Economics* 4.3 (1996): 364-370.

Bahmani-Oskooee, Mohsen, and A. B. M. Nasir. "Productivity bias hypothesis and the purchasing power parity: a review article." (2005).

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman. "Instrumental variables and GMM: Estimation and testing." *Stata journal* 3.1 (2003): 1-31.

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman. "ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression." *Boston College Department of Economics, Statistical Software Components S* 425401 (2007): 2007.

Belke, Ansgar, Frauke Dobnik, and Christian Dreger. "Energy consumption and economic growth: New insights into the cointegration relationshi " *Energy Economics* 33.5 (2011): 782-789.

Blackburne, Edward F., and Mark W. Frank. "Estimation of nonstationary heterogeneous panels." *Stata Journal* 7.2 (2007): 197.

Bound, John, David A. Jaeger, and Regina M. Baker. "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak." *Journal of the American statistical association* 90.430 (1995): 443-450.

BP (2017). *BP Energy Outlook - summary tables.* [online] London: BP. Available at: https://www.bp.com/en/global/corporate/energy-economics/energy-outlook.html [Accessed 3 Dec. 2017].

Brunnschweiler, Christa N., and Erwin H. Bulte. "The resource curse revisited and revised: A tale of paradoxes and red herrings." *Journal of environmental economics and management* 55.3 (2008): 248-264.

Camarero, Mariam. "The real exchange rate of the dollar for a panel of OECD countries: Balassa–Samuelson or distribution sector effect?." *Journal of Comparative Economics* 36.4 (2008): 620-632.

Chen, Yu-chin, and Kenneth Rogoff. "Commodity currencies and empirical exchange rate puzzles." *International Monetary Fund (IMF) Working Paper* 2 (2002): 27.

Chinn, Menzie David. *Sectoral productivity, government spending and real exchange rates: empirical evidence for OECD countries*. No. w6017. National Bureau of Economic Research, 1997.

Chinn, Menzie D. "The usual suspects? Productivity and demand shocks and Asia–Pacific real exchange rates." *Review of International Economics* 8.1 (2000): 20-43.

Chong, Yanping, Òscar Jordà, and Alan M. Taylor. "The Harrod–Balassa–Samuelson Hypothesis: Real Exchange Rates And Their Long-Run Equilibrium." *International Economic Review* 53.2 (2012): 609-634.

Choudhri, Ehsan U., and Mohsin S. Khan. "Real exchange rates in developing countries: Are Balassa-Samuelson effects present?." *IMF Staff Papers* (2005): 387-409.

Clague, Christopher. "Determinants of the national price level: some empirical results." *The Review of Economics and statistics* (1986): 320-323.

Clague, Christopher K. "Purchasing-Power Parities and Exchange Rates in Latin America." *Economic Development and Cultural Change* 36.3 (1988): 529-541.

Cragg, John G., and Stephen G. Donald. "Testing identifiability and specification in instrumental variable models." *Econometric Theory* 9.02 (1993): 222-240.

Damette, Olivier, and Majda Seghir. "Energy as a driver of growth in oil exporting countries?." *Energy Economics* 37 (2013): 193-199.

Davidson, Russell, and James G. MacKinnon. *Econometric theory and methods*. Vol. 5. New York: Oxford University Press, 2004.

De Gregorio, Jose, Alberto Giovannini, and Holger C. Wolf. "International evidence on tradables and nontradables inflation." *European Economic Review* 38.6 (1994): 1225-1244.

Devereux, Michael B. "Real Exchange Rates and the Balassa-Samuelson Effect Revisited." *NBER Reporter* 2014.4 (2014): 16.

De Vries, Margaret G. "Exchange Depreciation in Developing Countries." *Staff Papers* 15.3 (1968): 560-578.

Drine, I., and C. Rault. "How sure are we about the Balassa-Samuelson hypothesis? Time Series versus Panel Data Approach for Asian countries." *EUREQua, Sorbonne University and CNRS* (2002)

Dufour, Jean-Marie. "Some impossibility theorems in econometrics with applications to structural and dynamic models." *Econometrica: Journal of the Econometric Society* (1997): 1365-1387.

Dufour, Jean-Marie, and Joann Jasiak. "Finite sample limited information inference methods for structural equations and models with generated regressors." *International Economic Review* 42.3 (2001): 815-844.

Eden, S. H., and Been-Kwei Hwang. "The relationship between energy and GNP: further results." *Energy economics* 6.3 (1984): 186-190.

Egert, Balázs. "Equilibrium exchange rates in South Eastern Europe, Russia, Ukraine and Turkey: Healthy or (Dutch) diseased?." *Economic Systems* 29.2 (2005): 205-241.

Egert, Balázs, et al. "The Balassa–Samuelson effect in Central and Eastern Europe: myth or reality?." *Journal of comparative Economics* 31.3 (2003): 552-572.

Egert, Balázs, László Halpern, and Ronald MacDonald. "Equilibrium exchange rates in transition economies: taking stock of the issues." *Journal of Economic surveys* 20.2 (2006): 257-324.

Egert, Balazs, and Carol S. Leonard. "Dutch disease scare in Kazakhstan: Is it real?." *Open Economies Review* 19.2 (2008): 147-165.

Elliot, B. E., T. J. Rothenberg, and J. H. Stock. "Efficient tests of the unit root hypothesis." *Econometrica* 64.8 (1996): 13-36.

Engle, Robert F., and Clive WJ Granger. "Co-integration and error correction: representation, estimation, and testing." *Econometrica: journal of the Econometric Society* (1987): 251-276.

Faria, Joao Ricardo, and Miguel Leon-Ledesma. "Testing the Balassa–Samuelson effect: Implications for growth and the PPP." *Journal of Macroeconomics* 25.2 (2003): 241-253.

Fisher, R. A. "Statistical Methods for Research Workers; Oliver, Boyd, eds." (1932).

Frankel, Jeffrey. *The Currency-Plus-Commodity Basket; A Proposal for Exchange Rates in Oil-Exporting Countries to Accommodate Trade Shocks Automatically*. No. 333. Center for International Development at Harvard University, 2017.

García-Solanes, José, F. Israel Sancho-Portero, and Fernando Torrejón-Flores. "Beyond the Balassa–Samuelson effect in some new member states of the European Union." *Economic Systems* 32.1 (2008): 17-32.

García-Solanes, José, and Fernando Torrejón-Flores. "The Balassa-Samuelson hypothesis in developed countries and emerging market economies: different outcomes explained." *Economics: The Open-Access, Open-Assessment E-Journal* 3 (2009): 2.

Gubler, Matthias, and Christoph Sax. *The Balassa-Samuelson Effect Reversed: New Evidence from OECD Countries*. Faculty of Business and Economics-University of Basel, 2011.

Habib, Maurizio Michael, and Margarita M. Kalamova. "Are there oil currencies? The real exchange rate of oil exporting countries." (2007).

Hadri, Kaddour. "Testing for stationarity in heterogeneous panel data." *The Econometrics Journal* 3.2 (2000): 148-161.

Hahn, Jinyong, and Jerry Hausman. "A new specification test for the validity of instrumental variables." *Econometrica* 70.1 (2002): 163-189.

Hall, Alastair R., Glenn D. Rudebusch, and David W. Wilcox. "Judging instrument relevance in instrumental variables estimation." *International Economic Review* (1996): 283-298.

Harris, Richard DF, and Elias Tzavalis. "Inference for unit roots in dynamic panels where the time dimension is fixed." *Journal of econometrics* 91.2 (1999): 201-226.

Harrod, Roy F. International Economics, London: Cambridge University Press, 1933.

MacDonald, Ronald, and Luca Antonio Ricci. "Real exchange rates, imperfect substitutability, and imperfect competition." *Journal of Macroeconomics* 29.4 (2007): 639-664.

Hsieh, David A. "The determination of the real exchange rate: The productivity approach." *Journal of International Economics* 12.3-4 (1982): 355-362.

Huang, Bwo-Nung, Ming-Jeng Hwang, and Chin Wei Yang. "Causal relationship between energy consumption and GDP growth revisited: a dynamic panel data approach." *Ecological economics* 67.1 (2008): 41-54.

Im, Kyung So, M. Hashem Pesaran, and Yongcheol Shin. "Testing for unit roots in heterogeneous panels." *Journal of econometrics* 115.1 (2003): 53-74.

International Energy Agency (2015): World Energy Balances: Indicators, 1960-2014 (Data downloaded: 15 October 2015). UK Data Service,

International Energy Agency (2016): Indices of real and nominal end-use energy prices, 1978-2016 (Data downloaded: 15 September 2016). UK Data Service,

Kao, Chihwa, and Min-Hsien Chiang. "On the estimation and inference of a cointegrated regression in panel data." *Nonstationary panels, panel cointegration, and dynamic panels*. Emerald Group Publishing Limited, 2001. 179-222.

Kennedy, Peter. *A guide to econometrics*. MIT press, 2003.

Kilian, Lutz. "The economic effects of energy price shocks." *Journal of Economic Literature* 46.4 (2008): 871-909.

Kim, Byung-Yeon, and Iikka Korhonen. "Equilibrium exchange rates in transition countries: Evidence from dynamic heterogeneous panel models." *Economic Systems* 29.2 (2005): 144-162.

Kleibergen, Frank. "Pivotal statistics for testing structural parameters in instrumental variables regression." *Econometrica* 70.5 (2002): 1781-1803.

Kleibergen, Frank, and Richard Paap. "Generalized reduced rank tests using the singular value decomposition." *Journal of econometrics* 133.1 (2006): 97-126.

Korhonen, Iikka, and Tuuli Juurikkala. "Equilibrium exchange rates in oil-exporting countries." *Journal of Economics and Finance* 33.1 (2009): 71-79.

Kraft, John, and Arthur Kraft. "Relationship between energy and GNP." *J. Energy Dev.;(United States)* 3.2 (1978).

Lee, Chien-Chiang. "The causality relationship between energy consumption and GDP in G-11 countries revisited." *Energy Policy* 34.9 (2006): 1086-1093.

Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu. "Unit root tests in panel data: asymptotic and finite-sample properties." *Journal of econometrics* 108.1 (2002): 1-24.

MacDonald, Ronald, and Luca Antonio Ricci. "Real exchange rates, imperfect substitutability, and imperfect competition." *Journal of Macroeconomics* 29.4 (2007): 639-664.

MacDonald, Ronald, and Flavio V. Vieira. "A panel data investigation of real exchange rate misalignment and growth." (2010).

MacKinnon, James G. *Critical values for cointegration tests*. No. 1227. Queen's Economics Department Working Paper, 2010.

Maddala, Gangadharrao S., and Shaowen Wu. "A comparative study of unit root tests with panel data and a new simple test." *Oxford Bulletin of Economics and statistics* 61.S1 (1999): 631-652.

Mariano, Roberto S. "The existence of moments of the ordinary least squares and two-stage least squares estimators." *Econometrica: Journal of the Econometric Society* (1972): 643-652.

Mark, Nelson C., and Donggyu Sul. "Cointegration vector estimation by panel DOLS and long-run money demand." *Oxford Bulletin of Economics and Statistics* 65.5 (2003): 655-680.

Masih, Abul MM, and Rumi Masih. "On the temporal causal relationship between energy consumption, real income, and prices: some new evidence from Asian-energy dependent NICs based on a multivariate cointegration/vector error-correction approach." *Journal of policy modeling* 19.4 (1997): 417-440.

Mehlum, Halvor, Karl Moene, and Ragnar Torvik. "Institutions and the resource curse." *The economic journal* 116.508 (2006): 1-20.

Mikusheva, Anna. "Survey on statistical inferences in weakly-identified instrumental variable models." *Applied Econometrics* 29.1 (2013): 117-131.

Mohammadi, Hassan, and Shahrokh Parvaresh. "Energy consumption and output: Evidence from a panel of 14 oil-exporting countries." *Energy Economics* 41 (2014): 41-46.

Moreira, Marcelo J. "A conditional likelihood ratio test for structural models." *Econometrica* 71.4 (2003): 1027-1048.

Nagar, A. L. "The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations." *Econometrica: Journal of the Econometric Society* (1959): 575-595.

Nelson, Charles, and Richard Startz. "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument Is a Poor One." *The Journal of Business* 63.1 (1990): S125-40.

Nelson, Charles R., and Richard Startz. "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator." *Econometrica* 58.4 (1990): 967-76.

Njindan Iyke, Bernard. "The Penn effect revisited: new evidence from Latin America." *Review of Development Economics* 21.4 (2017): 1364-1379.

Ng, Serena, and Pierre Perron. "Lag length selection and the construction of unit root tests with good size and power." *Econometrica* 69.6 (2001): 1519-1554.

Ng, Serena, and Pierre Perron. "Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag." *Journal of the American Statistical Association* 90.429 (1995): 268-281.

Olea, José Luis Montiel, and Carolin Pflueger. "A robust test for weak instruments." *Journal of Business & Economic Statistics* 31.3 (2013): 358-369.

Officer, Lawrence H. "The productivity bias in purchasing power parity: An econometric investigation." *Staff Papers* 23.3 (1976): 545-579.

Oomes, Nienke, and Katerina Kalcheva. "Diagnosing Dutch disease: does Russia have the symptoms?." (2007).

Ozturk, Ilhan. "A literature survey on energy–growth nexus." *Energy policy* 38.1 (2010): 340-349.

Ozturk, Ilhan, Alper Aslan, and Huseyin Kalyoncu. "Energy consumption and economic growth relationship: Evidence from panel data for low and middle income countries." *Energy Policy* 38.8 (2010): 4422-4428.

Pedroni, Peter. "Critical values for cointegration tests in heterogeneous panels with multiple regressors." *Oxford Bulletin of Economics and statistics* 61.s 1 (1999): 653-670.

Pedroni, Peter. "Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis." *Econometric theory* 20.03 (2004): 597-625.

Pesaran, M. Hashem. "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica* 74.4 (2006): 967-1012.

Pesaran, M. Hashem. "On the interpretation of panel unit root tests." *Economics Letters* 116.3 (2012): 545-546.

Pesaran, M. Hashem. "Testing weak cross-sectional dependence in large panels." *Econometric Reviews* 34.6-10 (2015): 1089-1117.

Pesaran, M. Hashem, Yongcheol Shin, and Ron P. Smith. "Pooled mean group estimation of dynamic heterogeneous panels." *Journal of the American Statistical Association* 94.446 (1999): 621-634.

Pesaran, M. Hashem, and Ron Smith. "Estimating long-run relationships from dynamic heterogeneous panels." *Journal of econometrics* 68.1 (1995): 79-113.

Reese, Simon, and Joakim Westerlund. "Panicca: Panic on Cross-Section Averages." *Journal of Applied Econometrics* (2015).

Rickne, Johanna. *Oil prices and real exchange rate movements in oil-exporting countries: The role of institutions*. No. 810. IFN Working Paper, 2009.

Rodrik, Dani. "The real exchange rate and economic growth." *Brookings papers on economic activity* 2008.2 (2008): 365-412.

Rogoff, Kenneth. "Traded Goods Consumption Smoothing and the Random Walk Behavior of the Real Exchange Rate." *Monetary and Economic Studies* 10.2 (1992): 1-29.

Ross, Michael. *The oil curse: How petroleum wealth shapes the development of nations*. Princeton University Press, 2012.

Rothenberg, Thomas J. "Approximating the distributions of econometric estimators and test statistics." *Handbook of econometrics* 2 (1984): 881-935.

Sachs, Jeffrey D., and Andrew M. Warner. *Natural resource abundance and economic growth*. No. w5398. National Bureau of Economic Research, 1995.

Sachs, Jeffrey D., and Andrew M. Warner. "Fundamental sources of long-run growth." *The American economic review* 87.2 (1997): 184-188.

Sachs, Jeffrey D., and Andrew M. Warner. "The big push, natural resource booms and growth." *Journal of development economics* 59.1 (1999): 43-76.

Sachs, Jeffrey D., and Andrew M. Warner. "The curse of natural resources." *European economic review* 45.4 (2001): 827-838.

Samuelson, Paul A. "Theoretical notes on trade problems." *The Review of Economics and Statistics* (1964): 145-154.

Schwert, G. William. "Tests for Unit Roots: A Monte Carlo Investigation." *Journal of Business & Economic Statistics* (1989): 147-159.

Shea, John. "Instrument relevance in multivariate linear models: A simple measure." *Review of Economics and Statistics* 79.2 (1997): 348-352.

Soytas, Ugur, and Ramazan Sari. "Energy consumption and GDP: causality relationship in G-7 countries and emerging markets." *Energy economics* 25.1 (2003): 33-37.

Stern, David I. "A multivariate cointegration analysis of the role of energy in the US macroeconomy." *Energy Economics* 22.2 (2000): 267-283.

Staiger, Douglas, and James H. Stock. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65.3 (1997): 557-586.

Stock, James H., and Mark W. Watson. "A simple estimator of cointegrating vectors in higher order integrated systems." *Econometrica: Journal of the Econometric Society* (1993): 783-820.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. "A survey of weak instruments and weak identification in generalized method of moments." *Journal of Business & Economic Statistics* 20.4 (2002): 518-529.

Stock, James H., and Motohiro Yogo. "Testing for Weak Instruments in Linear IV Regression." *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (2005): 80.

Streimikiene, Dalia, and Rafał Kasperowicz. "Review of economic growth and energy consumption: A panel cointegration analysis for EU countries." *Renewable and Sustainable Energy Reviews* 59 (2016): 1545-1549.

Subramanian, Mr Arvind, and Xavier Sala-i-Martin. *Addressing the natural resource curse: an illustration from Nigeria*. No. 3-139. International Monetary Fund, 2003.

"The Atlas of Economic Complexity," Center for International Development at Harvard University, http://www.atlas.cid.harvard.edu

Van der Ploeg, Frederick. "Natural resources: Curse or blessing?." *Journal of Economic Literature* 49.2 (2011): 366-420.

Westerlund, Joakim. "Testing for error correction in panel data." *Oxford Bulletin of Economics and statistics* 69.6 (2007): 709-748.

Zalduendo, Mr Juan. *Determinants of Venezuela's equilibrium real exchange rate*. No. 6-74. International Monetary Fund, 2000.